

STIC-Biotech/ChemLib

154239

mgj

From: Walicka, Malgorzata
Sent: Monday, May 23, 2005 5:45 PM
To: STIC-Biotech/ChemLib

Please search SEQ ID NO:1, 3, 4, 5, 6 in the application No. 10/622,893.

Thank you in advance.

Malgorzata A. Walicka, Ph.D.
Patent Examiner
Art Unit 1652, Recombinant Enzymes
USPTO, Remsen Building, Room 2C76
400 Dulany St.
Alexandria, VA 22313
Mail Room 2C70
Tel. (571) 272-0944, fax (571) 273-0944

RECEIVED
MAY 24 2005
(STIC)

STAFF USE ONLY

Searcher: _____
Searcher Phone: 2-_____
Date Searcher Picked up: 5/27/05
Date Completed: 5/27/05
Searcher Prep/Rev. Time: _____
Online Time: _____

Type of Search

NA#: 1 AA#: 4
Interference: _____ SPDI: _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure#: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable

STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: auto
WWW/Internet: _____
Other(Specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:47:36 ; Search time 3.47974 Seconds
(without alignments)
1189.648 Million cell updates/sec

Title: US-10-622-893A-1

Perfect score: 63

Sequence: 1 MGGSGDDDDLAL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1462099 seqs, 344972447 residues

Total number of hits satisfying chosen parameters: 1462099

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	100.0	12	17	US-10-622-893A-1
2	63	100.0	472	17	US-10-622-893A-5
3	49	77.8	102	16	US-10-767-701-46722
4	49	77.8	149	16	US-10-767-701-46955
5	48	76.2	64	14	US-10-080-170-503
6	48	76.2	64	16	US-10-080-170-503
7	48	76.2	64	16	US-10-468-356-503
8	48	76.2	230	16	US-10-767-701-40998
9	47	74.6	173	16	US-10-437-963-200561
10	47	74.6	230	16	US-10-437-963-172039
11	46	73.0	103	14	US-10-289-454-386
12	46	73.0	625	16	US-10-437-963-152224
13	46	73.0	926	16	US-10-437-963-174460

14	69.8	75	16	US-10-437-963-133890	Sequence 133890,
15	69.8	359	16	US-10-437-963-116167	Sequence 116167,
16	69.8	652	16	US-10-437-963-198055	Sequence 198055,
17	69.8	757	16	US-10-437-963-181509	Sequence 181509,
18	68.3	143	16	US-10-767-701-32972	Sequence 32972, A
19	68.3	209	15	US-10-425-114-62751	Sequence 62751, A
20	68.3	312	16	US-10-437-963-182122	Sequence 182122,
21	68.3	344	16	US-10-437-963-141563	Sequence 141563,
22	68.3	361	16	US-10-437-963-146049	Sequence 146049,
23	68.3	361	16	US-10-437-963-182133	Sequence 182133,
24	68.3	477	15	US-10-425-114-63251	Sequence 63251, A
25	68.3	646	16	US-10-437-963-161859	Sequence 161859,
26	66.7	15	15	US-10-323-069A-136	Sequence 136, App
27	66.7	15	17	US-10-851-965-136	Sequence 136, App
28	66.7	16	14	US-10-010-160-61	Sequence 61, Appl
29	66.7	17	14	US-10-010-160-66	Sequence 66, Appl
30	66.7	57	16	US-10-437-963-144342	Sequence 144342,
31	66.7	80	16	US-10-767-701-48120	Sequence 48120, A
32	66.7	125	16	US-10-437-963-111374	Sequence 111374,
33	66.7	157	15	US-10-259-194A-120	Sequence 120, App
34	66.7	178	16	US-10-437-963-104580	Sequence 104580,
35	66.7	223	15	US-10-424-599-245391	Sequence 245391,
36	66.7	232	15	US-10-424-599-193534	Sequence 193534,
37	66.7	247	16	US-10-437-963-129496	Sequence 129496,
38	66.7	275	16	US-10-437-963-107388	Sequence 107388,
39	66.7	341	16	US-10-437-963-107394	Sequence 107394,
40	66.7	404	15	US-10-282-122A-63165	Sequence 63165, A
41	66.7	417	17	US-10-732-923-13830	Sequence 13830, A
42	66.7	421	16	US-10-437-963-126589	Sequence 126589,
43	66.7	447	15	US-10-374-780A-1140	Sequence 1140, Ap
44	66.7	447	15	US-10-412-699B-1338	Sequence 1338, Ap
45	66.7	447	16	US-10-437-963-110675	Sequence 110675,

ALIGNMENTS

RESULT 1
US-10-622-893A-1
; Sequence 1, Application US/10622893A
; Publication No. US20050014935A1
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Sheng
; APPLICANT: Datta, Abhijit
; APPLICANT: Wang, Yuping
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DETERMINATION OF GLYCATED PROTEINS
; FILE REFERENCE: 466992001300
; CURRENT APPLICATION NUMBER: US/10/622,893A
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 40%-100% identity to leader sequence
US-10-622-893A-1

Query Match 100.0%; Score 63; DB 17; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGGSGDDDDLAL 12
| | | | | | | | | | | | | |
Db 1 MGGSGDDDDLAL 12

RESULT 2
US-10-622-893A-5
; Sequence 5, Application US/10622893A
; Publication No. US20050014935A1

```
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Sheng
; APPLICANT: Datta, Abhijit
; APPLICANT: Wang, Yuping
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DETERMINATION OF GLYCATED PROTEINS
; FILE REFERENCE: 466992001300
; CURRENT APPLICATION NUMBER: US/10/622,893A
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric protein
US-10-622-893A-5

Query Match          100.0%; Score 63; DB 17; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.061;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGGSGDDDDLAL 12
Db 1 MGGSGDDDDLAL 12
|||||
|

RESULT 3
US-10-767-701-46722
; Sequence 46722, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 46722
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 7554236.pep
US-10-767-701-46722

Query Match          77.8%; Score 49; DB 16; Length 102;
Best Local Similarity 90.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGSGDDDDLAL 11
Db 74 GGGSGDDDDAA 83
|||||
|

RESULT 4
US-10-767-701-46955
; Sequence 46955, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
```

```
; SEQ ID NO 46955
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C49_289.pep
US-10-767-701-46955

Query Match          77.8%; Score 49; DB 16; Length 149;
Best Local Similarity 90.0%; Pred. No. 3.3;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGSGDDDDLAL 11
Db 91 GGGSGDDDDAA 100
|||||
|

RESULT 5
US-10-080-170-503
; Sequence 503, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 503
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-503

Query Match          76.2%; Score 48; DB 14; Length 64;
Best Local Similarity 80.0%; Pred. No. 2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGSGDDDDLAL 11
Db 10 GGGSGDDDDLAL 19
|||||
|

RESULT 6
US-10-080-170-503
; Sequence 503, Application US/10080170
; Publication No. US20040121322A9
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 503
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-503

Query Match          76.2%; Score 48; DB 16; Length 64;
Best Local Similarity 80.0%; Pred. No. 2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

QY 2 GSGGDDDDLA 11
|||:|
Db 10 GGGGDDDDIA 19

RESULT 7

US-10-468-356-503
; Sequence 503, Application US/10468356
; Publication No. US20040197896A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENEOMICS AS A TOOL FOR
; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 05394.0019
; CURRENT APPLICATION NUMBER: US/10/468,356
; CURRENT FILING DATE: 2003-08-19
; PRIOR APPLICATION NUMBER: 10/080,170
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 655
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 503
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-468-356-503

Query Match 76.2%; Score 48; DB 16; Length 64;
Best Local Similarity 80.0%; Pred. No. 2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGGDDDDLA 11
|||:|
Db 10 GGGGDDDDIA 19

RESULT 8

US-10-767-701-40998
; Sequence 40998, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 40998
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C15329_1.pep
US-10-767-701-40998

Query Match 76.2%; Score 48; DB 16; Length 230;
Best Local Similarity 66.7%; Pred. No. 7.4;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGGGDDDDLA 12
:|:|:|
Db 15 LGGAGDDDLAI 26

RESULT 9

US-10-437-963-200561
; Sequence 200561, Application US/10437963

; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 200561
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_96018C.1.pep
US-10-437-963-200561

Query Match 74.6%; Score 47; DB 16; Length 173;
Best Local Similarity 72.7%; Pred. No. 8;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSGGDDDDLA 12
|||:|
Db 154 GGDGDDDDAAM 164

RESULT 10

US-10-437-963-172039
; Sequence 172039, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 172039
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_70212C.1.pep
US-10-437-963-172039

Query Match 74.6%; Score 47; DB 16; Length 230;
Best Local Similarity 80.0%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGGDDDDLA 11
|||:|
Db 93 GGDGDDDDLS 102

RESULT 11

US-10-289-454-386
; Sequence 386, Application US/10289454
; Publication No. US20030157479A1

; GENERAL INFORMATION:
; APPLICANT: Bachmann, Martin
; APPLICANT: Jennings, Gary
; APPLICANT: Sonderesger, Ivo
; TITLE OF INVENTION: Antigen Arrays for Treatments of Allergic Eosinophilic Diseases
; FILE REFERENCE: 1700.0360001
; CURRENT APPLICATION NUMBER: US/10/289,454
; CURRENT FILING DATE: 2003-02-10
; PRIOR APPLICATION NUMBER: US 60/396,636
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: PCT/IB02/00166
; PRIOR FILING DATE: 2002-01-21
; PRIOR APPLICATION NUMBER: US 10/050,902
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/331,045
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 396
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 386
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Mouse His-eotaxin-C1
US-10-289-454-386

Query Match 73.0%; Score 46; DB 14; Length 103;
Best Local Similarity 90.0%; Pred. No. 6.8;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSGDDDDLAL 12
Db 10 GSGDDDDKAL 19

RESULT 12
US-10-437-963-152224
; Sequence 152224, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 152224
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_52297C.1.pep
US-10-437-963-152224

Query Match 73.0%; Score 46; DB 16; Length 625;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGDDDD 9
Db 144 GSGDDDD 151

RESULT 13
US-10-437-963-174460
; Sequence 174460, Application US/10437963
; Publication No. US20040123343A1

; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 174460
; LENGTH: 926
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_72399C.1.pep
US-10-437-963-174460

Query Match 73.0%; Score 46; DB 16; Length 926;
Best Local Similarity 88.9%; Pred. No. 66;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGDDDDL 10
Db 666 GSGDDDDL 674

RESULT 14
US-10-437-963-133890
; Sequence 133890, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 133890
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_35717C.1.pep
US-10-437-963-133890

Query Match 69.8%; Score 44; DB 16; Length 75;
Best Local Similarity 66.7%; Pred. No. 10;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSAGDDDRVEL 12
Db 1 MGSAGDDDRVEL 12

RESULT 15
US-10-437-963-116167
; Sequence 116167, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 116167
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_19694C.1.pep
US-10-437-963-116167

Query Match 69.8%; Score 44; DB 16; Length 359;
Best Local Similarity 66.7%; Pred. No. 52;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGGSGDDDDLAL 12
:|:|:|:|:
Db 44 VGGGEDGDLAL 55

Search completed: May 27, 2005, 14:15:01
Job time : 4.47974 secs

THIS PAGE BLANK (USPTO)

3/21/2003

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:28:05 ; Search time 4.55437 Seconds
(without alignments)
1019.048 Million cell updates/sec

Title: US-10-622-893A-1
Perfect score: 63
Sequence: 1 MCGSGDDDLAL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	76.2	64	5	ABU05852 M. tuberc
2	46	73.0	96	7	ADK17286 Virus-lik
3	43	68.3	130	5	ABJ05339 Human PTH
4	43	68.3	300	8	ADQ36935 Cell prol
5	43	68.3	300	8	ADQ15603 Rice stre
6	43	68.3	378	6	AAQ19572 M sterili
7	42	66.7	15	7	ADH52757 pBAD-TOPO
8	42	66.7	17	5	ABG68924 Lawsonia
9	42	66.7	15	5	ABG68919 Lawsonia
10	42	66.7	15	8	ADI45189 Rice isop
11	42	66.7	402	4	ABG12558 Novel hum
12	42	66.7	404	6	ABU5241 Protein e
13	42	66.7	409	8	ADL05698 M. catarr
14	42	66.7	447	8	ADI42677 Plant tra
15	42	66.7	447	8	ADO02925 Thalecres
16	42	66.7	581	4	ABBS5983 Drosophil
17	42	66.7	730	4	ABB61435 Drosophil
18	42	66.7	784	4	ABB62115 Drosophil
19	42	66.7	858	8	ADS20230 Human agg
20	42	66.7	1138	7	ABO77206 Pseudomon
21	42	66.7	1926	4	AAG84915 Shrimp wh
22	41.5	65.9	21	5	AAU75830 Wheat xyl
23	41	65.1	32	5	ABG80719 Human Igg
24	41	65.1	32	8	ADI40815 pCep-BK-F
25	41	65.1	47	5	ABG80720 Human IgG

26	41	65.1	47	8	ADI40817 pCep-SP-E
27	41	65.1	61	4	AAU22386 Human car
28	41	65.1	61	7	ADE46354 Human car
29	41	65.1	61	8	ADJ07772 Human col
30	41	65.1	67	3	AAB53618 Human col
31	41	65.1	103	5	ABG94342 Human res
32	41	65.1	103	5	ABG80654 Human res
33	41	65.1	107	5	ABG94322 Mouse res
34	41	65.1	107	5	ABG80634 Processed
35	41	65.1	124	5	ABG94340 Mouse mPr
36	41	65.1	124	5	ABG80652 Mouse trp
37	41	65.1	124	7	ADD24200 mPrPt-EK-
38	41	65.1	130	5	ABG80722 Mouse Res
39	41	65.1	130	8	ADI40821 Resistin-
40	41	65.1	185	8	ADR09188 Human pro
41	41	65.1	216	7	ADJ68201 Human hea
42	41	65.1	350	5	ABG94339 Mouse mPr
43	41	65.1	350	5	ABG80651 Mouse pri
44	41	65.1	350	7	ADD24199 mPrPt-EK-
45	41	65.1	357	7	ADJ70832 Human hea

ALIGNMENTS

RESULT 1
ABU05852
ID ABU05852 standard; protein; 64 AA.
XX
AC ABU05852;
XX
DT 08-APR-2003 (first entry)
XX
DE M. tuberculosis and M. leprae marker protein #503.
XX
KW Mycobacterioses; survival; virulence; protective antigen; vaccine;
KW Mycobacterial disease; tuberculosis; leprosy.
XX
OS Mycobacterium tuberculosis.
OS Mycobacterium leprae.
PN W0200274903-A2.
XX
PD 26-SEP-2002.
XX
PF 22-FEB-2002; 2002WO-IB001973.
XX
PR 22-FEB-2001; 2001US-0270123P.
XX (INSP) INST PASTEUR.
XX Cole S;
XX WFI; 2002-759895/82.
XX
PT Identifying and selecting genes for survival or virulence of mycobacteria
PT by a comparative genomic analysis of the sequences of Mycobacterium
PT tuberculosis and M. leprae.
XX
PS Claim 17; Page 714-715; 874pp; English.
XX
CC This invention relates to a novel method for identifying essential genes
CC for survival or virulence of mycobacteria species. The method comprises
CC aligning the genomic sequence of a first mycobacterium species on a
CC genomic sequence of a second mycobacterium species and selecting a
CC polynucleotide sequence that is highly conserved in both genomes with no
CC counterparts in other bacterial genomic sequences and that corresponds to
CC an essential gene for the survival or virulence of mycobacterium species.
CC The method of the invention is useful for detecting M. tuberculosis or M.
CC leprae infection. The method reduces the number of potential new targets
CC and protective antigens for new drugs and vaccine compositions to treat
CC and prevent mycobacterial diseases, particularly tuberculosis and
CC leprosy. The present sequence represents a marker protein from

CC Mycobacterium tuberculosis and Mycobacterium leprae identified using the
CC method of the invention
XX
SQ Sequence 64 AA;

Query Match 76.2%; Score 48; DB 5; Length 64;
Best Local Similarity 80.0%; Fred. No. 3.6;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGGDDDDLA 11
|||
DB 10 GGGGDDDDIA 19
|||

RESULT 2
ADK17286
ID ADK17286 standard; peptide; 96 AA.
XX ADK17286;
XX
XX 06-MAY-2004 (first entry)
XX
XX Virus-like particle repetitive antigen array peptide.
DE
XX
XX antiallergic; antiasthmatic; cytostatic; vaccine; virus-like particle;
KW interleukin; IL-5; IL-13; ectaxin; repetitive antigen array;
KW allergic eosinophilic disease; asthma; Hodgkin's lymphoma.
XX
XX Unidentified.
OS
XX
XX WO2003040164-A2.
FN
XX
XX 15-MAY-2003.
PD
XX
XX 07-NOV-2002; 2002WO-EP012455.
PF
XX
XX 07-NOV-2001; 2001US-0331045P.
PR
XX 18-JAN-2002; 2002US-00050902.
PR
XX 21-JAN-2002; 2002WO-IB000166.
PR
XX 19-JUL-2002; 2002US-0356636P.
XX
XX (CYTO-) CYTOS BIOTECHNOLOGY AG.
PA
XX
XX Bachmann M, Jennings G, Sonderegger I;
PI
XX
XX WPI; 2003-441518/41.
DR
XX
XX Composition comprising an ordered and repetitive antigen or antigenic
PT determinant array, useful as a medicament, or for manufacturing a
PT medicament for treating allergic eosinophilic diseases, e.g. asthma, or
PT Hodgkin's lymphoma.
PT
XX
XX . Disclosure; SEQ ID NO 400; 245pp; English.
PS
XX
XX The invention relates to a composition comprising a virus-like particle
CC and at least one antigen, which is a protein or peptide of interleukin
CC (IL)-5, IL-13 or ectaxin and is bound to the virus-like particle, or a
CC core particle with at least one first attachment site and at least one
CC antigen with at least one second attachment site, where the antigen is a
CC protein or peptide of IL-5, IL-13 or ectaxin. The second attachment site
CC is an attachment site naturally or not naturally occurring with the
CC antigen or antigenic determinant. The second attachment site is capable
CC of association to the first attachment site, and where the antigen or
CC antigenic determinant and the core particle interact through the
CC association to form an ordered and repetitive antigen array. The
CC compositions are useful as medicaments, or for manufacturing a medicament
CC or a vaccine for treating allergic eosinophilic diseases, e.g. asthma, or
CC Hodgkin's lymphoma and related diseases. This sequence is used to
CC generate the compound of the invention.
XX
XX Sequence 96 AA;
SQ

Query Match 73.0%; Score 46; DB 7; Length 96;

```

XX DE Cell proliferation-related polypeptide #45.
XX KW cell proliferation related polypeptide; cell proliferation; senescence;
XX KM differentiation; stress response.
XX OS Oryza sativa.
XX XX WO2004061122-A2.
XX PN 22-JUL-2004.
XX PD 23-DEC-2003; 2003WO-US041200.
XX PF 26-DEC-2002; 2002US-0436565P.
XX PR (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PA Cooper B;
XX PI WPI; 2004-534388/51.
XX DR New nucleic acid molecule encoding a cell proliferation-related
XX PT polypeptide, useful for modulating cell proliferation, senescence,
XX PT differentiation, development, and stress response in plants, and for
XX PT producing enhanced food crops.
XX SQ Claim 28; SEQ ID NO 94; 408pp; English.
XX
XX The present invention relates to an isolated nucleic acid molecule
XX CC encoding a cell proliferation-related polypeptide. The nucleic acid
XX CC molecule and the encoded polypeptide, and methods are useful for
XX CC modulating cell proliferation, senescence, differentiation, development,
XX CC and stress response in plants, and for producing enhanced food crops. The
XX CC present sequence represents a cell proliferation-related polypeptide. The
XX CC present sequence is published separately from the main body of the
XX CC specification as EPO data.
XX SQ Sequence 300 AA;
XX
XX Query Match 68.3%; Score 43; DB 8; Length 300;
XX Best Local Similarity 66.7%; Pred. No. 1.1e+02;
XX Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 MGGSGDDDDIAL 12
DB :|||:|||
46 IGGGGEDDDAL 57

RESULT 5
ADQ15603
ID ADQ15603 standard; protein; 300 AA.
XX AC ADQ15603;
XX XX 07-OCT-2004 (first entry)
XX DE Rice stress-related protein #6.
XX KW rice; stress-related protein; plant maturation; plant development;
XX KW plant proliferation; plant senescence; plant disease-resistance;
XX KW plant stress response; transgenic plant; pest tolerance;
XX KW herbicide tolerance; biotic stress tolerance; abiotic stress tolerance;
XX KW improved nutritional value; increased yield; increased proliferation.
XX OS Oryza sativa.
XX XX WO2004061080-A2.
XX PN 22-JUL-2004.
XX PD 23-DEC-2003; 2003WO-US041098.
XX PF
XX XX

XX DE Cell proliferation-related polypeptide #45.
XX KW cell proliferation related polypeptide; cell proliferation; senescence;
XX KM differentiation; stress response.
XX OS Oryza sativa.
XX XX WO2004061122-A2.
XX PN 22-JUL-2004.
XX PD 23-DEC-2003; 2003WO-US041200.
XX PF 26-DEC-2002; 2002US-0436565P.
XX PR (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PA Cooper B;
XX PI WPI; 2004-534388/51.
XX DR New nucleic acid molecule encoding a cell proliferation-related
XX PT polypeptide, useful for modulating cell proliferation, senescence,
XX PT differentiation, development, and stress response in plants, and for
XX PT producing enhanced food crops.
XX SQ Claim 28; SEQ ID NO 94; 408pp; English.
XX
XX The present invention relates to an isolated nucleic acid molecule
XX CC encoding a cell proliferation-related polypeptide. The nucleic acid
XX CC molecule and the encoded polypeptide, and methods are useful for
XX CC modulating cell proliferation, senescence, differentiation, development,
XX CC and stress response in plants, and for producing enhanced food crops. The
XX CC present sequence represents a cell proliferation-related polypeptide. The
XX CC present sequence is published separately from the main body of the
XX CC specification as EPO data.
XX SQ Sequence 300 AA;
XX
XX Query Match 68.3%; Score 43; DB 8; Length 300;
XX Best Local Similarity 66.7%; Pred. No. 1.1e+02;
XX Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 MGGSGDDDDIAL 12
DB :|||:|||
46 IGGGGEDDDAL 57

RESULT 6
AAO19572
ID AAO19572 standard; protein; 378 AA.
XX AC AAO19572;
XX XX 28-JAN-2003 (first entry)
XX DT M sterilia protein fragment #7.
XX DE
XX KW Substance PF1022; biosynthesis; transgenic; chorismic acid; anthelmintic;
XX KW phenylalanine; p-aminophenylpyruvic acid; pharmaceutical;
XX KW veterinary drug.
XX OS Mycelia sterilia.
XX XX WO200277244-A1.
XX PN 03-OCT-2002.
XX PD
XX PF 22-MAR-2002; 2002WO-JP002782.
XX XX 22-MAR-2001; 2001JP-00082227.
XX PR (MEIJ ) MEIJI SEIKA KAISHA LTD.
XX PA Yanai K, Sumida N, Watanabe M, Moriya T, Murakami T;
XX PI WPI; 2003-018934/01.
XX DR
XX XX Novel biosynthesis gene-transferred transformants for producing PF1022
XX PT substance derivatives by fermentation, as pharmaceuticals or veterinary
XX PT drugs with anthelmintic activity.
XX XX Claim 24; Page 109-111; 116pp; Japanese.

```

XX The present invention relates to transformants capable of producing
 CC PF1022 substance derivatives. These were obtained by transferring a gene
 CC participating in the biosynthesis pathway from chorismic acid to p-
 CC aminophenylpyruvic acid (biosynthesis gene) into a phenylalanine-
 CC requiring host derived from an organism producing the PF1022 substance.
 CC The transformants are producing PF1022 substance derivatives by
 CC fermentation, for use as pharmaceuticals or veterinary drugs. The present
 CC sequence is a fragment of a Mycelia sterilia protein described in the
 CC exemplification of the invention
 XX
 SQ Sequence 378 AA;

Query Match 68.3%; Score 43; DB 6; Length 378;
 Best Local Similarity 80.0%; Pred. No. 1.4e+02;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GSGDDDDLLAL 12
 DB 238 GEGDDDDAAL 247
 |||||

RESULT 7
 ADH52757
 ID ADH52757 standard; peptide; 15 AA.

XX ADH52757;
 AC
 DT 25-MAR-2004 (first entry)
 XX
 DE pBAD-TOPO:Oprf vector amino-terminal peptide.
 XX
 KW pigmented anaerobic bacterium; 16S rRNA; antiinflammatory; vaccine;
 KW Periodontal disease; Oprf; pBAD-TOPO:Oprf vector.
 XX
 OS Synthetic.

XX WO2003054755-A2.

PN 03-JUL-2003.

XX 19-DEC-2002; 2002WO-IB005539.

XX 21-DEC-2001; 2001US-0342999P.

XX (PFIZ) PFIZER PROD INC.

XX Hardham JM, King KW, Krishnan R, McGavin DR, Dreier KJ;

XX WPI; 2003-559192/52.

PT New isolated pigmented anaerobic bacteria, useful for preparing a vaccine
 PT for treating or preventing periodontal disease in companion animals.

XX Example; SEQ ID NO 136; 159pp; English.

XX The invention relates to a novel isolated pigmented anaerobic bacteria
 CC having a 16S rRNA DNA sequence fully defined in the specification,
 CC provided that the bacteria is not a strain of Porphyromonas gingivalis
 CC designated as dog 20B. The invention has antiinflammatory applications
 CC whilst the bacteria and polynucleotide may be useful for preparing a
 CC vaccine for treating or preventing periodontal disease in companion
 CC animals. The current sequence is that of the Porphyromonas-related vector
 CC amino-terminal peptide of the invention.

XX Sequence 15 AA;

Query Match 66.7%; Score 42; DB 7; Length 15;
 Best Local Similarity 85.7%; Pred. No. 6.9;
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 MG-GSGDDDD-LAL 12
 |||||

Db 1 MGS GSGDDDDKLAL 14

RESULT 8
 AEG68924

ID AEG68924 standard; peptide; 17 AA.

XX AEG68924;

DT 10-OCT-2002 (first entry)

DE Lawsonia intracellularis immunogenic protein related peptide #6.

XX Immunogen; fliH; fliR; ntrC; glnH; motA; motB; tlyC; ytfM; ytfN; porcine;
 KW pig; avian; bird; porcine proliferative enteropathy; PPE;
 KW intestinal adenomatosis complex; porcine intestinal adenomatosis; PIA;
 KW necrotic enteritis; proliferative haemorrhagic enteropathy; PHE;
 KW regional ileitis; haemorrhagic bowel syndrome; vaccine; antibacterial;
 KW porcine proliferative enteritis; Campylobacter spp.-induced enteritis.

OS Synthetic.

XX WO200238594-A1.

PN 16-MAY-2002.

XX 09-NOV-2001; 2001WO-AU001462.

XX 10-NOV-2000; 2000AU-00001381.

XX 17-NOV-2000; 2000US-0249596P.

XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

XX (AUPO-) AUSTRALIAN PORK LTD.

XX (PFIZ) PFIZER PROD INC.

PI Rosey EL, King KW, Good RT, Strugnell RA;

XX WPI; 2002-557448/59.

XX New immunogenic polypeptide comprising epitope of Lawsonia spp.
 PT polypeptide such as fliH, fliR, ntrC, glnH, motA, polypeptides, useful in
 PT vaccines for treatment of porcine proliferative enteropathy in pigs and
 PT birds.

XX Example 2; Page 153; 155pp; English.

XX The invention describes an isolated or recombinant immunogenic
 CC polypeptide (I) which comprises, mimics or cross-reacts with a B-cell or
 CC T-cell epitope of a Lawsonia spp. polypeptide such as fliH, fliR, ntrC,
 CC glnH, motA, motB, tlyC, ytfM or ytfN polypeptides. (I) is useful for
 CC identifying whether or not a porcine or avian animal has suffered from a
 CC past infection, or is currently infected, with Lawsonia spp or a
 CC microorganism that is immunologically cross-reactive with Lawsonia spp.
 CC Antibodies are useful for diagnosing infection of a porcine or avian
 CC animal by Lawsonia spp. or a microorganism that is immunologically cross-
 CC reactive with Lawsonia spp. A nucleic acid encoding a Lawsonia spp.
 CC immunogen is useful as probes or primers for detecting Lawsonia spp. or
 CC related microorganism in a biological sample derived from a porcine or
 CC avian animal subject. (I) is preferably useful for vaccinating porcine
 CC animals against intestinal diseases collectively known as porcine
 CC proliferative enteropathy (PPE), previously known as intestinal
 CC adenomatosis complex, porcine intestinal adenomatosis (PIA), necrotic
 CC enteritis, proliferative haemorrhagic enteropathy, regional ileitis,
 CC haemorrhagic bowel syndrome, porcine proliferative enteritis and
 CC Campylobacter spp.-induced enteritis. (I) is also useful in vaccines for
 CC the prophylaxis and treatment of PPE in birds. This sequence represents a
 CC Lawsonia intracellularis immunogenic protein related peptide

XX Sequence 17 AA;

Query Match 66.7%; Score 42; DB 5; Length 17;
 Best Local Similarity 85.7%; Pred. No. 7.9;
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 MG-GSGDDDD-LAL 12
 DB 1 MGSGSGDDDDKLAL 14

RESULT 9
 ABG68919
 ID ABG68919 standard; peptide; 17 AA.
 XX
 AC ABG68919;
 XX
 DT 10-OCT-2002 (first entry)
 XX
 DE Lawsonia intracellularis immunogenic protein related peptide #1.
 XX
 KW Immunogen; fihB; flir; ntrC; glnH; motA; motB; tlyC; ytfM; ytfN; porcine;
 KW pig; avian; bird; porcine proliferative enteropathy; PPE;
 KW intestinal adenomatosis complex; porcine intestinal adenomatosis; PIA;
 KW necrotic enteritis; proliferative haemorrhagic enteropathy;
 KW regional ileitis; haemorrhagic bowel syndrome; vaccine; antibacterial;
 KW porcine proliferative enteritis; Campylobacter spp.-induced enteritis.
 XX
 OS Synthetic.
 XX
 PN WO200238594-A1.
 XX
 PD 16-MAY-2002.
 XX
 XX 09-NOV-2001; 2001WO-AU001462.
 XX
 PR 10-NOV-2000; 2000AU-00001381.
 PR 17-NOV-2000; 2000US-0249596P.
 XX
 PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
 PA (AUPO-) AUSTRALIAN PORK LTD.
 PA (PFIZ) PFIZER PROD INC.
 XX
 PI Rosey EL, King KW, Good RT, Strugnell RA;
 XX
 DR WPI; 2002-557448/59.
 XX
 XX New immunogenic polypeptide comprising epitope of Lawsonia spp.
 PT polypeptide such as fihB, flir, ntrC, glnH, motA, polypeptides, useful in
 PT vaccines for treatment of porcine proliferative enteropathy in pigs and
 PT birds.
 XX
 PS Example 2; Page 152; 155pp; English.
 XX
 CC The invention describes an isolated or recombinant immunogenic
 CC polypeptide (I) which comprises, mimics or cross-reacts with a B-cell or
 CC T-cell epitope of a Lawsonia spp. polypeptide such as fihB, flir, ntrC,
 CC glnH, motA, motB, tlyC, ytfM or ytfN polypeptides. (I) is useful for
 CC identifying whether or not a porcine or avian animal has suffered from a
 CC past infection, or is currently infected, with Lawsonia spp. or a
 CC microorganism that is immunologically cross-reactive with Lawsonia spp.
 CC Antibodies are useful for diagnosing infection of a porcine or avian
 CC animal by Lawsonia spp. or a microorganism that is immunologically cross-
 CC reactive with Lawsonia spp. A nucleic acid encoding a Lawsonia spp.
 CC immunogen is useful as probes or primers for detecting Lawsonia spp. or
 CC related microorganism in a biological sample derived from a porcine or
 CC avian animal subject. (I) is preferably useful for vaccinating porcine
 CC animals against intestinal diseases collectively known as porcine
 CC proliferative enteropathy (PPE), previously known as intestinal
 CC adenomatosis complex, porcine intestinal adenomatosis (PIA), necrotic
 CC enteritis, proliferative haemorrhagic enteropathy, regional ileitis,
 CC haemorrhagic bowel syndrome, porcine proliferative enteritis and
 CC Campylobacter spp.-induced enteritis. (I) is also useful in vaccines for
 CC the prophylaxis and treatment of PPE in birds. This sequence represents a
 CC Lawsonia intracellularis immunogenic protein related peptide
 XX
 SQ Sequence 17 AA;

Query Match 66.7%; Score 42; DB 5; Length 17;
 Best Local Similarity 85.7%; Pred. No. 7.9;
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 MG-GSGDDDD-LAL 12
 DB 1 MGSGSGDDDDKLAL 14

RESULT 10
 ADI45189
 ID ADI45189 standard; protein; 157 AA.
 XX
 AC ADI45189;
 XX
 DT 22-APR-2004 (first entry)
 XX
 DE Rice isoprenoid biosynthesis-associated protein #60.
 XX
 KW Rice; isoprenoid biosynthesis; plant; isopentenyl diphosphate; IPP;
 KW dimethylallyl alcohol; DMAPP; short-chain plastid prenyltransferase;
 KW gibberellin; carotenoid; abscisic acid; tocopherol; plastoquinone;
 KW phyloquinone; mevalonate pathway; phytosterol; brassinosteroid;
 KW ubiquinone; monoterpene; sesquiterpene; protein prenylation; chlorophyll;
 KW haeme; yield.
 XX
 OS Oryza sativa.
 XX
 PN US2004010815-A1.
 XX
 PD 15-JAN-2004.
 XX
 XX 26-SEP-2002; 2002US-00259194.
 XX
 PR 26-SEP-2001; 2001US-0325277P.
 PR 04-APR-2002; 2002US-0370620P.
 PR 04-APR-2002; 2002US-0370743P.
 XX
 PA (LANG/) LANGE B M.
 PA (GHAS/) GHASSEMIAN M.
 PA (BRIG/) BRIGGS S P.
 PA (COOP/) COOPER B.
 PA (GLAZ/) GLAZEBROOK J.
 PA (GOFF/) GOFF S A.
 PA (KATA/) KATAGIRI F.
 PA (KREP/) KREPS J.
 PA (MOUG/) MOUGHAMER T.
 PA (PROV/) PROVART N.
 PA (RICK/) RICHE D.
 PA (ZHUT/) ZHU T.
 XX
 PI Lange BM, Ghassenian M, Briggs SP, Cooper B, Glazebrook J;
 PI Goff SA, Katagiri F, Kreps J, Moughamer T, Provart N, Ricke D;
 PI Zhu T;
 XX
 DR WPI; 2004-090562/09.
 DR N-PSDB; ADI45188.
 XX
 XX New isolated polynucleotides and polypeptides associated with isoprenoid
 PT synthesis in plants, useful for producing transgenic plants, for targeted
 PT gene disruption, as well as markers or probes.
 PT
 XX
 PS Claim 4; SEQ ID NO 120; 117pp; English.
 XX
 CC The invention relates to a polynucleotide (or its complement, protein
 CC encoding fragment or reverse complement), comprising a nucleotide
 CC sequence encoding a polypeptide comprising an amino acid sequence
 CC involved in or associated with the biosynthesis of isoprenoids in a rice
 CC plant. Also included are an isolated polypeptide involved in or
 CC associated with the biosynthesis of isoprenoids in a plant, an expression
 CC cassette comprising the polynucleotide, a host cell comprising the
 CC expression cassette, and a transgenic plant comprising the expression
 CC cassette. The polypeptides and polynucleotides include those associated

CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX SQ Sequence 404 AA;

Query Match 66.7%; Score 42; DB 6; Length 404;
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGGDDDD 9
 || |||||
 Db 381 GGGGDDDD 388

RESULT 13
 ADL05698
 ID ADL05698 standard; protein; 409 AA.

AC ADL05698;
 XX
 DT 06-MAY-2004 (first entry)
 DE
 DE M. catarrhalis protein #1464.
 XX Moraxella catarrhalis; infection.
 XX Moraxella catarrhalis.

OS
 XX US6673910-B1.
 XX
 XX 06-JAN-2004.
 XX
 XX 04-APR-2000; 2000US-00540236.
 XX
 XX 08-APR-1999; 99US-0128416P.
 XX
 XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton GL;

XX WPI; 2004-178127/17.
 DR N-PSDB; ADL03778.

XX New nucleic acid encoding a Moraxella catarrhalis polypeptide, useful for
 PT preparing a composition for diagnosing, preventing or treating infection

PT caused by Moraxella catarrhalis.
 XX
 PS Disclosure; SEQ ID NO 384; 429pp; English.

XX The invention relates to an isolated nucleic acid encoding an Moraxella
 CC catarrhalis polypeptide. The nucleic acid is useful for preparing a
 CC composition for diagnosing, preventing or treating infection caused by
 CC Moraxella catarrhalis. The present sequence represents the amino acid
 CC sequence of a M. catarrhalis protein.

XX SQ Sequence 409 AA;

Query Match 66.7%; Score 42; DB 8; Length 409;
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGGDDDD 9
 || |||||
 Db 386 GGGGDDDD 393

RESULT 14
 ADI42677
 ID ADI42677 standard; protein; 447 AA.

XX AC ADI42677;
 XX
 DT 22-APR-2004 (first entry)
 XX
 DE Plant transcription factor #408.

XX transgenic; plant; enhanced tolerance to abiotic stress;
 KW phosphate tolerance; hormone sensitivity; disease resistance;
 KW sugar sensing; flowering; flower structure; stem bifurcation;
 KW branching pattern; apical dominance; trichome; stem morphology;
 KW root growth; root hair; seed development; cell proliferation;
 KW cell differentiation; premature senescence; necrosis; plant size;
 KW leaf morphology; seed morphology; seed biochemistry; root anthocyanin;
 KW plant anthocyanin; light response; shade avoidance; bioinformatic;
 KW transcription factor; gene; ds.

OS Oryza sativa.

XX
 PN US2004019927-A1.

XX
 PD 29-JAN-2004.

XX
 PF 25-FEB-2003; 2003US-00374780.

XX
 PR 18-APR-2001; 2001US-00837944.

XX (SHER/) SHERMAN B K.
 PA (RIEC/) RIECHMANN J L.
 PA (JIAN/) JIANG C.
 PA (HEAR/) HEARD J E.
 PA (HAAR/) HAAKE V.
 PA (CREE/) CREELMAN R A.
 PA (RATC/) RATCLIFFE O.
 PA (ADAM/) ADAM L J.
 PA (REUB/) REUBER T L.
 PA (KEDD/) KEDDIE J.
 PA (BROU/) BROUN P E.
 PA (PILG/) PILGRIM M L.
 PA (DUBE/) DUBELL A N.
 PA (PINE/) PINEDA O.
 PA (YUGG/) YU G.

XX Sherman BK, Riechmann JL, Jiang C, Heard JE, Haake V;
 PI Creelman RA, Ratcliffe O, Adam LJ, Reuber TL, Keddie J, Broun PE;
 PI Pilgrim ML, Dubell AN, Pineda O, Yu G;

XX WPI; 2004-132245/13.

XX

PT New transgenic plant comprising a recombinant polynucleotide of any one
PT of more than 500 nucleotide sequences, useful in bioinformatic search
PT methods.

XX Claim 1; SEQ ID NO 1140; 435pp; English.

CC The invention describes a transgenic plant comprising a recombinant
CC polynucleotide of any one of more than 500 nucleotide sequences fully
CC defined in the specification or its complement. The method of the
CC invention can be used to produced a plant having altered traits such as:
CC enhanced tolerance to abiotic stress; glyphosphate tolerance; hormone
CC sensitivity; disease resistance; sugar sensing; early or late flowering;
CC altered flower structure, change in stem bifurcations, altered branching
CC pattern; reduced apical dominance, reduced trichome density; lack of
CC trichomes; reduced ectopic trichome development; altered trichome
CC development; increase in trichome number; altered stem morphology;
CC increased root growth; increased root hairs; altered seed development;
CC altered cell proliferation or cell differentiation; rapid development;
CC premature senescence; increased necrosis; increase in seedling or plant
CC size; decreased plant size; leaf morphology; seed morphology; seed
CC biochemistry; increase in root anthocyanins; increase in plant
CC anthocyanins, or alteration in light response or shade avoidance. The
CC transgenic plant, polynucleotides and polypeptides are useful in
CC bioinformatic search methods. This is the amino acid sequence of a plant
CC transcription factor, and an orthologue of Arabidopsis thaliana
CC transcription factors isolated in the invention, that can be used in the
CC creation of a transgenic plant with altered traits.

XX Sequence 447 AA;

Query Match 66.7%; Score 42; DB 8; Length 447;

Best Local Similarity 72.7%; Pred. No. 2.4e+02;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGSGDDDLAL 12

DB 75 GGGGDDDLGL 85

RESULT 15

AD002925
ID ADO02925 standard; protein; 447 AA.

XX ADO02925;

DT 01-JUL-2004 (first entry)

XX Thalecress transcription factor, Rice orthologue #66.

XX Rice; transcription factor; plant; transgenic; abiotic stress;
KW cold tolerance; heat tolerance; drought; osmotic stress;
KW phosphate limitation; potassium limitation; nitrogen limitation;
KW hormone sensitivity; disease resistance; sugar sensing; seed germination;
KW flowering; inflorescence architectural change;
KW meristem cell differentiation; phyllotaxy; apical dominance;
KW trichome development; seed development; premature senescence;
KW delayed senescence; lethality; necrosis; plant size; leaf morphology;
KW seed morphology; secondary metabolism; light response; shade avoidance.

OS *Oryza sativa*.

XX US2004045049-A1.

PN 04-MAR-2004.

XX 10-APR-2003; 2003US-00412699.

XX 13-SEP-1999; 99US-00394519.

PR 21-JAN-2000; 2000US-00489176.

PR 17-FEB-2000; 2000US-00506720.

PR 22-MAR-2000; 2000US-00532591.

PR 22-MAR-2000; 2000US-00533029.

PR 22-MAR-2000; 2000US-00533030.

PR 22-MAR-2000; 2000US-00533392.
PR 22-MAR-2000; 2000US-00533648.
PR 06-APR-2000; 2000WO-US009448.
PR 16-NOV-2000; 2000US-00713994.
PR 27-MAR-2001; 2001US-00819142.
PR 17-APR-2001; 2001US-00837444.
PR 30-JAN-2002; 2002US-00958131.
PR 14-JUN-2002; 2002US-00171468.
PR 09-AUG-2002; 2002US-00225066.
PR 09-AUG-2002; 2002US-00225067.
PR 09-AUG-2002; 2002US-00225068.
PR 17-DEC-2002; 2002US-0434166P.
PR 25-FEB-2003; 2003US-00374780.

XX (ZHAN/) ZHANG J.
PA (FROM/) FROMM M E.
PA (HEAR/) HEARD J E.
PA (RIEC/) RIECHMANN J L.
PA (ADAM/) ADAM L J.
PA (BROU/) BROUN P E.
PA (PINE/) PINEDA O.
PA (REUB/) REUBER T L.
PA (KEDD/) KEDDIE J S.
PA (YUGG/) YU G.
PA (JIAN/) JIANG C.
PA (SAMA/) SAMAHA R S.
PA (PILG/) PILGRIM M L.
PA (CREE/) CREELMAN R A.
PA (DUBE/) DUBELL A N.
PA (RATC/) RATCLIFFE O.
PA (KUMI/) KUMIMOTO R.
PA (SHER/) SHERMAN B K.

PI Zhang J, Fromm ME, Heard JE, Riechmann JL, Adam LJ, Broun PE,
PI Pineda O, Reuber TL, Reddie JS, Yu G, Jiang C, Samaha RS;
PI Pilgrim ML, Creelman RA, Dubell AN, Ratcliffe O, Kumimoto R;
PI Sherman BK;

XX WPI; 2004-225755/21.

PT New transgenic plant, useful in developing phenotypes with altered or
PT improved characteristics or traits.

XX Claim 1; SEQ ID NO 1339; 213pp; English.

XX The invention relates to a transgenic plant comprises a recombinant
CC polynucleotide having a polynucleotide sequence or its complementary
CC sequence comprising a sequence encoding a polypeptide, that initiates
CC transcription (i.e. a transcription factor) from Arabidopsis, Soybean,
CC Rice, Rape or Corn, comprising any of the sequences appearing as ADO01588
CC -ADO03527 or ADO03530-ADO03559. Also included are using a transgenic
CC plant to grow a progeny plant, an expression cassette (comprising a
CC constitutive, inducible or tissue-specific promoter and a recombinant
CC polynucleotide described above), a host cell comprising the expression
CC cassette, producing a modified plant having a modified trait, identifying
CC a factor that is modulated by or interacts with a polypeptide encoded by
CC the polynucleotide sequence and identifying at least one downstream
CC polynucleotide sequence that is subject to a regulatory effect of any of
CC the polypeptides encoded by the polynucleotide described above. The
CC transgenic plant is useful for producing a plant that has an altered
CC trait e.g. an enhanced tolerance to abiotic stress (increased tolerance
CC to chilling, germination in cold conditions, freezing tolerance, tolerance
CC to heat, tolerance to drought, tolerance to osmotic stress, tolerance to
CC salt, tolerance to phosphate limitation, tolerance to potassium
CC limitation, decreased sensitivity to nitrogen limitation), altered
CC hormone sensitivity, reduced sensitivity to abscisic acid, an altered
CC response to ethylene, disease resistance, altered susceptibility to
CC Botrytis, altered susceptibility to Fusarium, altered susceptibility to
CC Erysiphe, altered susceptibility to Pseudomonas syringae, altered
CC susceptibility to Sclerotinia, altered sugar sensing, improved seed
CC germination and seedling vigor, early flowering, late flowering, extended
CC period of flowering, an inflorescence architectural change, a change in
CC stem bifurcations, a lack of a shoot meristem, reduced meristem cell

CC differentiation, altered phyllotaxy, altered branching pattern, reduced
 CC apical dominance, reduced trichome density, ectopic trichome development,
 CC altered trichome development, altered stem morphology, increased root
 CC growth, increased root hairs, altered seed development, altered cell
 CC proliferation/cell differentiation, premature senescence, delayed
 CC senescence, lethality, increased necrosis, an increase in seedling or
 CC plant size, decreased plant size, a change in leaf morphology, increased
 CC altered leaf development, increased leaf size and mass, glossy leaves,
 CC leaf cell expansion, change in seed morphology, altered seed coloration,
 CC increased seed size, decreased seed size, altered seed shape, change in
 CC leaf biochemistry, increased leaf wax, an alteration in leaf prenyl lipid
 CC content, increased leaf insoluble sugars, decreased leaf insoluble
 CC sugars, increased leaf anthocyanins, an alteration of leaf fatty acid
 CC content, an alteration of leaf glucosinolate content, change in seed
 CC biochemistry, an increase in seed oil content, decrease in seed oil
 CC content, increase in seed fatty acid content, decrease in seed fatty acid
 CC content, increase in seed protein content, decrease in seed protein
 CC content, alteration in seed prenyl lipid content, increase in seed
 CC sterols, upregulation of genes involved in secondary metabolism, increase
 CC in root anthocyanins, increase in plant anthocyanins, and alteration in
 CC light response or shade avoidance. The present sequence represents an
 CC orthologue of a thalecress transcription factor isolated from Rice.
 XX
 SQ Sequence 447 AA;

Query Match 66.7%; Score 42; DB 8; Length 447;
 Best Local Similarity 72.7%; Pred. No. 2.4e+02;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGSDDDDLAL 12
 |||||
 Db 75 GGGDDDLAL 85

Search completed: May 27, 2005, 13:43:51
 Job time : 7.55437 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:34:21 ; Search time 0.921109 Seconds
(without alignments)
1253.491 Million cell updates/sec

Title: US-10-622-893A-1

Perfect score: 63

Sequence: 1 MGSGDDDDLLAL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 79:**

2: PIR1:**

3: PIR2:**

4: PIR3:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	76.2	64	2 B70512	hypothetical prote
2	44	69.8	240	2 H84708	hypothetical prote
3	43	68.3	894	2 T27007	hypothetical prote
4	42	66.7	585	2 S46825	hypothetical prote
5	42	66.7	730	2 A53064	folded gastrulatio
6	42	66.7	818	2 T02823	probable membrane
7	42	66.7	918	2 S44769	C29E4.2 protein -
8	42	66.7	943	2 S54493	hypothetical prote
9	42	66.7	1360	2 T18403	asparagine/asparta
10	41	65.1	84	2 B87376	hypothetical prote
11	41	65.1	489	2 T36100	probable ATP-bind
12	41	65.1	508	1 ISH000	protein disulfide-
13	41	65.1	510	1 ISB050	protein disulfide-
14	41	65.1	815	2 T46169	CDC48-like protein
15	41	65.1	875	1 S66672	phosphatidylinosit
16	40	63.5	251	1 RNBV3C	DNA-directed RNA p
17	40	63.5	294	2 C82643	daunorubicin C-13
18	40	63.5	334	2 A29561	prostatic spermine
19	40	63.5	353	2 E95914	probable secreted
20	40	63.5	504	2 G84251	hypothetical prote
21	40	63.5	539	2 G95405	hypothetical prote
22	40	63.5	825	2 JS0174	cellulase [EC 3.2.
23	40	63.5	1074	2 T52654	Ca2+-transporting
24	40	63.5	1150	2 D88556	protein B0464.2 [i
25	40	63.5	1274	2 S28279	hypothetical prote
26	40	63.5	6420	2 T30283	polyketide synthas
27	39	61.9	58	2 S03091	Band 3 anion trans
28	39	61.9	185	2 T30047	hypothetical prote
29	39	61.9	397	2 H70753	hypothetical prote

30	39	61.9	407	2 T51543	TOM (target of myb
31	39	61.9	509	1 ISMSSS	protein disulfide-
32	39	61.9	579	2 S69069	hypothetical prote
33	39	61.9	598	2 F83508	flagellar M-ring p
34	39	61.9	633	2 D83091	ATP sulfurylase GT
35	39	61.9	633	2 T14612	hypothetical prote
36	39	61.9	810	2 T19879	hypothetical prote
37	38	60.3	63	2 C87074	conserved hypothet
38	38	60.3	157	2 S09814	hypothetical prote
39	38	60.3	218	2 D83161	hypothetical prote
40	38	60.3	288	2 B40722	homeotic protein 1
41	38	60.3	360	2 B84376	hypothetical prote
42	38	60.3	476	1 C70986	probable serine/th
43	38	60.3	508	1 ISRTSS	protein disulfide-
44	38	60.3	523	2 A71246	lysine-tRNA ligase
45	38	60.3	539	2 T23700	hypothetical prote

ALIGNMENTS

RESULT 1

B70512

hypothetical protein Rv2111c - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004

C;Accession: B70512

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: B70512

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-64 <COL>

A;Cross-references: UNIPROT:O33246; GB:297559; GB:AL123456; NID:g3261820; PIDN:CAB10702.

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: Rv2111c

Query Match 76.2%; Score 48; DB 2; Length 64;
Best Local Similarity 80.0%; Pred. No. 0.42;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy	2	GGSGDDDDLA	11
Db	10	GGGGDDDDIA	19

RESULT 2

H84708

hypothetical protein At2g30480 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004

C;Accession: H84708

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.W.; Venter, J.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: H84708

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-240 <STO>

A;Cross-references: UNIPROT:O23686; GB:AE002093; NID:g2257712; PIDN:AB863096.1; GSPDB:GN

C;Genetics:

A;Gene: At2g30480

A;Map position: 2

```
Query Match      69.8%; Score 44; DB 2; Length 240;
Best Local Similarity 77.8%; Pred. No. 7.4;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 MGGSGDDDD 9
Db      163 VGGTGGDDDD 171
      :||:|||||
      :||:|||||

RESULT 3
T27007
hypothetical protein Y48B6A.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T27007
R:Wall, M.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z20297
A:Accession: T27007
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-894 <WIL>
A:Cross-references: UNIPROT:Q9U297; EMBL:AL110490; NID:el542263; PIDN:CAB54451.1; CESP:Y48B6A
A:Experimental source: clone Y48B6A
C:Genetics:
A:Gene: CESP:Y48B6A.11
A:Introns: 71/2; 146/2; 510/1; 558/2

Query Match      68.3%; Score 43; DB 2; Length 894;
Best Local Similarity 77.8%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 GGGSGDDDDL 10
Db      609 GGGGDDDDV 617
      |||:|||||
      |||:|||||

RESULT 4
S46825
hypothetical protein YHL010c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jul-2004
C:Accession: S46825
R:Favella, T.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid 9433.
A:Reference number: S46796
A:Accession: S46825
A:Molecule type: DNA
A:Residues: 1-585 <FAV>
A:Cross-references: UNIPROT:P38748; EMBL:U11582; NID:g2289793; PID:g508747; GSPDB:GN000000000
C:Genetics:
A:Gene: MIPS:YHL010c
A:Cross-references: SGD:S0001002
A:Map position: 8L
C:Superfamily: fission yeast probable zinc finger protein SPAC16E8.13; RING finger homol
F:236-285/Domain: RING finger homology <RNG>

Query Match      66.7%; Score 42; DB 2; Length 585;
Best Local Similarity 77.8%; Pred. No. 40;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 MGGSGDDDD 9
Db      373 VGGSGDDDN 381
      :||:|||||
      :||:|||||

RESULT 5
A53064
folded gastrulation (fog) precursor - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C:Accession: A53064
```

```
R:Costa, M.; Wilson, E.T.; Wieschaus, E.
Cell 76, 1075-1089, 1994
A:Title: A putative cell signal encoded by the folded gastrulation gene coordinates cell
A:Reference number: A53064; MUID:94185167; PMID:8137424
A:Accession: A53064
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-730 <COS>
A:Cross-references: UNIPROT:P40795; GB:U03717; NID:g430722; PIDN:AAA18955.1; PID:g430723
C:Genetics:
A:Gene: fog
A:Cross-references: FlyBase:FBgn0000719

Query Match      66.7%; Score 42; DB 2; Length 730;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 MGGSGDDDD 9
Db      513 IGGRGDDDD 521
      :|||:|||||
      :|||:|||||

RESULT 6
T02823
probable membrane protein L1231.1 [imported] - Leishmania major (strain Friedlin)
C:Species: Leishmania major
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C:Accession: C81459; T02823
R:Myler, P.J.; Audleman, L.; deVos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.; F
Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999
A:Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c
A:Reference number: A81455; MUID:99178987; PMID:10077609
A:Accession: C81459
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-818 <PYL>
A:Cross-references: UNIPROT:Q25353; GB:AE001274; NID:g3264850; PIDN:AAC24646.1; PID:g3264
A:Experimental source: strain MHOM/IL/81/Friedlin
C:Genetics:
A:Gene: L1231.1
A:Map position: 1
C:Superfamily: Leishmania major probable membrane protein L1231.1
C:Keywords: transmembrane protein

Query Match      66.7%; Score 42; DB 2; Length 818;
Best Local Similarity 87.5%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GGGSGDDDD 9
Db      711 GGGGDDDD 718
      |||:|||||
      |||:|||||

RESULT 7
S44769
C29E4.2 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 23-Mar-2001
C:Accession: S44769
R:Wilson, R.
submitted to the EMBL Data Library, September 1993
A:Description: Sequence of the C. elegans cosmid C29E4.
A:Reference number: S44738
A:Accession: S44769
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-918 <WIL>
A:Cross-references: EMBL:L23651; NID:g388575; PID:g388580
C:Genetics:
A:Introns: 73/2; 97/3; 131/3; 381/3; 557/1; 684/1; 769/1; 830/3
C:Superfamily: Caenorhabditis elegans C29E4.2 protein

Query Match      66.7%; Score 42; DB 2; Length 918;
```

Best Local Similarity 87.5%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGGDDDD 9
|||
Db 512 GGGGDDDD 519

RESULT 8
S54493
hypothetical protein YMR124w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YMR564.06
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C;Accession: S54493; S48514
R;Lye, G.; Church, C.M.
A;Reference number: S54014
A;Accession: S54493
A;Molecule type: DNA
A;Residues: 1-943 <LYE>
A;Cross-references: UNIPROT:P39523; EMBL:Z49273; NID:G809577; PIDN:CAA89273.1; PID:G809577
R;Pandit, S.; Sternglanz, R.
submitted to the EMBL Data Library, May 1995
A;Reference number: S48514
A;Accession: S48514
A;Molecule type: DNA
A;Residues: 770-943 <PAN>
A;Cross-references: EMBL:L07650; NID:G349191; PIDN:AAA35122.1; PID:G349192
C;Genetics:
A;Cross-references: SGD:S0004731
A;Map position: 13R

Query Match 66.7%; Score 42; DB 2; Length 943;
Best Local Similarity 87.5%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGGDDDD 9
|||
Db 168 GGDGDDDD 175

RESULT 9
T18403
asparagine/aspartate rich protein - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18403
R;BARALE, J.C.; ATTALBONNEFOY, G.; BRAHIMI, K.; DASILVA, L.P.; LANGSLEY, G.
Mol. Biochem. Parasitol. 87, 169-181, 1997
A;Title: Plasmodium-falciparum asparagine and aspartate rich protein-2 is an evolutionary
A;Reference number: Z08834; MUID:97391121; PMID:9247928
A;Accession: T18403
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1360 <BAR>
A;Cross-references: UNIPROT:Q94649; EMBL:Y08924; NID:e1008106; PID:e1154303; PIDN:CAA701
C;Genetics:
A;Gene: aarp2

Query Match 66.7%; Score 42; DB 2; Length 1360;
Best Local Similarity 87.5%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGGDDDD 9
|||
Db 399 GGDGDDDD 406

RESULT 10
B87376
hypothetical protein CC1022 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus

Best Local Similarity 87.5%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGGDDDD 9
|||
Db 512 GGGGDDDD 519

RESULT 8
S54493
hypothetical protein YMR124w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YMR564.06
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C;Accession: S54493; S48514
R;Lye, G.; Church, C.M.
A;Reference number: S54014
A;Accession: S54493
A;Molecule type: DNA
A;Residues: 1-943 <LYE>
A;Cross-references: UNIPROT:P39523; EMBL:Z49273; NID:G809577; PIDN:CAA89273.1; PID:G809577
R;Pandit, S.; Sternglanz, R.
submitted to the EMBL Data Library, May 1995
A;Reference number: S48514
A;Accession: S48514
A;Molecule type: DNA
A;Residues: 770-943 <PAN>
A;Cross-references: EMBL:L07650; NID:G349191; PIDN:AAA35122.1; PID:G349192
C;Genetics:
A;Cross-references: SGD:S0004731
A;Map position: 13R

Query Match 66.7%; Score 42; DB 2; Length 943;
Best Local Similarity 87.5%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGGDDDD 9
|||
Db 168 GGDGDDDD 175

RESULT 9
T18403
asparagine/aspartate rich protein - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18403
R;BARALE, J.C.; ATTALBONNEFOY, G.; BRAHIMI, K.; DASILVA, L.P.; LANGSLEY, G.
Mol. Biochem. Parasitol. 87, 169-181, 1997
A;Title: Plasmodium-falciparum asparagine and aspartate rich protein-2 is an evolutionary
A;Reference number: Z08834; MUID:97391121; PMID:9247928
A;Accession: T18403
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1360 <BAR>
A;Cross-references: UNIPROT:Q94649; EMBL:Y08924; NID:e1008106; PID:e1154303; PIDN:CAA701
C;Genetics:
A;Gene: aarp2

Query Match 66.7%; Score 42; DB 2; Length 1360;
Best Local Similarity 87.5%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGGDDDD 9
|||
Db 399 GGDGDDDD 406

RESULT 10
B87376
hypothetical protein CC1022 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: B87376
R;Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87376
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84 <STO>
A;Cross-references: UNIPROT:Q9A9G6; GB:AE005673; NID:gl3422314; PIDN:AAK23006.1; GSPDB:G
C;Genetics:
A;Gene: CC1022

Query Match 65.1%; Score 41; DB 2; Length 84;
Best Local Similarity 70.0%; Pred. No. 7.5;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGGDDDDLA 11
|||
Db 35 GGGGDDDELA 44

RESULT 11
T36100
probable ATP-binding protein - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T36100
R;Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21596
A;Accession: T36100
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-489 <SAU>
A;Cross-references: UNIPROT:Q9X873; EMBL:AL049661; PIDN:CAB41216.1; GSPDB:GN00070; SCOD
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCODB:SCE134.18

Query Match 65.1%; Score 41; DB 2; Length 489;
Best Local Similarity 66.7%; Pred. No. 48;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MGSGGDDDDLA 12
:|||||
Db 323 LGGSGDDDTALVL 334

RESULT 12
ISHUSS
protein disulfide-isomerase (EC 5.3.4.1) precursor - human
N;Alternate names: cellular thyroid hormone-binding protein; endoplasmic reticulum prote
hydroxylase; proline,2-oxoglutarate 4-dioxygenase; prollyl 4-hydroxylase beta chain; pro
N;Contains: procollagen-proline dioxygenase (EC 1.14.11.2) beta chain; thyroxine deiodin
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1992 #sequence_revision 17-Nov-1995 #text_change 09-Jul-2004
C;Accession: A31913; B31913; A30055; A29787; A26632; A42791; I33178; F61002
R;Tasanen, K.; Parkkonen, T.; Chow, L.T.; Kivirikko, K.I.; Pihlajaniemi, T.
J. Biol. Chem. 263, 16218-16224, 1988
A;Title: Characterization of the human gene for a polypeptide that acts both as the beta
A;Reference number: A31913; MUID:89034087; PMID:2846539
A;Accession: A31913
A;Molecule type: DNA
A;Residues: 1-118;209-508 <TAS>
A;Cross-references: UNIPROT:P07237
A;Accession: B31913
A;Molecule type: DNA
A;Residues: 1-508 <TA2>
A;Cross-references: GB:J04049; GB:J04050; GB:M22803; GB:M22804; GB:M22805; GB:M22806; NI

R;Morris, J.I.; Varandani, P.T.
 Biochim. Biophys. Acta 949, 169-180, 1988
 A;Title: Characterization of a cDNA for human glutathione-insulin transhydrogenase (prot
 A;Reference number: A30055; MUID:88134925; PMID:3342239
 A;Accession: A30055
 A;Molecule type: mRNA
 A;Residues: 293-351; 'R', 353-401; 'HS', 404-459; 'Q', 461-508 <MOR>
 A;Cross-references: EMBL:X07077
 A;Experimental source: liver
 R;Cheng, S.; Gong, Q.; Parkison, C.; Robinson, E.A.; Appella, E.; Merlino, G.T.; Pastan,
 J. Biol. Chem. 262, 11221-11227, 1987
 A;Title: The nucleotide sequence of a human cellular thyroid hormone binding protein pre
 A;Reference number: A29787; MUID:87280213; PMID:3611107
 A;Accession: A29787
 A;Molecule type: mRNA
 A;Residues: 1-140; 'R', 142-359; 'RAG', 363-371; 'P', 373-508 <CH>
 A;Cross-references: GB:J02783; NID:G339646; PIDN:AAA61169.1; PID:G339647
 A;Note: parts of this sequence, including the amino end of the mature protein, were dete
 R;Pihlajaniemi, T.; Helaakoski, T.; Tasanen, K.; Myllyla, R.; Huhtala, M.L.; Koivu, J.;
 EMBO J. 6, 643-649, 1987
 A;Title: Molecular cloning of the beta-subunit of human prollyl 4-hydroxylase. This subun
 A;Reference number: A26632; MUID:87218523; PMID:3034602
 A;Accession: A26632
 A;Molecule type: mRNA
 A;Residues: 1-9; 'PMX', 13-43; 'PP', 46-48; 'H', 50-438; 'G', 440-443; 'G', 445-480; 'V', 482-508 <H
 A;Cross-references: EMBL:X05130; NID:G35654; PIDN:CAA28775.1; PID:G35655
 A;Note: parts of this sequence, including the amino end of the mature protein, were dete
 R;Tasanen, K.; Oikarinen, J.; Kivirikko, K.I.; Pihlajaniemi, T.
 J. Biol. Chem. 267, 11513-11519, 1992
 A;Title: Promoter of the gene for the multifunctional protein disulfide isomerase polype
 A;Reference number: A42791; MUID:92283868; PMID:1597478
 A;Accession: A42791
 A;Molecule type: DNA
 A;Residues: 1-25 <TA3>
 A;Cross-references: GB:S37207; NID:G249922; PIDN:AB22262.1; PID:G249923
 A;Note: sequence extracted from NCBI backbone (NCBIN:104755, NCBI:P:104756)
 R;Ward, L.D.; Hong, J.; Whitehead, R.H.; Simpson, R.J.
 Electrophoresis 11, 883-891, 1990
 A;Title: Development of a database of amino acid sequences for human colon carcinoma pro
 A;Reference number: A33178; MUID:91176935; PMID:2079031
 A;Accession: I33178
 A;Molecule type: protein
 A;Residues: 18-24; 'X', 26 <WAR>
 R;Bauw, G.; Rasmussen, H.H.; Van Den Bulcke, M.; Van Damme, J.; Puype, M.; Gesser, B.; C
 Electrophoresis 11, 528-536, 1990
 A;Title: Two-dimensional gel electrophoresis, protein electroblotting and microsequencin
 A;Reference number: A61002; MUID:91031404; PMID:1699755
 A;Accession: F61002
 A;Molecule type: protein
 A;Residues: 317-326; 351-363; 'X', 365-370; 402-406; 'X', 408-419 <BAU>
 C;Genetics:
 A;Gene: GDB:P4HB; PO4DB
 A;Cross-references: GDB:120708; OMIM:1176790
 A;Map position: 17q25-17q25
 A;Introns: 49/1; 118/1; 162/3; 208/3; 243/3; 285/3; 352/3; 393/1; 453/3; 482/3
 C;Complex: in procollagen-proline dioxygenase forms a heterotetramer of two alpha chains
 C;Function: <DSI>
 A;Description: as protein disulfide-isomerase (EC 5.3.4.1) catalyzes the rearrangement o
 A;Pathway: protein synthesis
 C;Function: <PHB>
 A;Description: as procollagen-proline dioxygenase (EC 1.14.11.2) beta chain participates
 A;Pathway: collagen synthesis
 C;Function: <TDI>
 A;Description: as thyroxine deiodinase (EC 3.8.1.4) reduces iodinated thyronine residues
 C;Superfamily: protein disulfide-isomerase; thioredoxin homology
 C;Keywords: ascorbic acid; duplication; endoplasmic reticulum; homodimer; intramolecular
 F;1-17/Domain: signal sequence #status predicted <SIG>
 F;18-508/Product: protein disulfide-isomerase #status predicted <MAT>
 F;32-119/Domain: thioredoxin homology <TX1>
 F;375-460/Domain: thioredoxin homology <TX2>
 F;505-508/Region: endoplasmic reticulum retention signal
 F;53-56; 397-400/Disulfide bonds: redox-active #status predicted

F;312-343/Disulfide bonds: #status predicted
 Query Match 65.1%; Score 41; DB 1; Length 508;
 Best Local Similarity 87.5%; Pred. No. 50;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 3 GSGDDDDL 10
 Db 477 GAGDDDDL 484
 |:|||||
 RESULT 13
 ISBOSS
 protein disulfide-isomerase (EC 5.3.4.1) precursor - bovine
 N;Alternate names: cellular thyroid hormone-binding protein; endoplasmic reticulum protei
 N;Contains: procollagen-proline dioxygenase (EC 1.14.11.2) beta chain; thyroxine deiodina
 C;Species: Bos primigenius taurus (cattle)
 C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
 C;Accession: A26829
 R;Yamauchi, K.; Yamamoto, T.; Hayashi, H.; Koya, S.; Takikawa, H.; Toyoshima, K.; Horiuc
 Biochem. Biophys. Res. Commun. 146, 1485-1492, 1987
 A;Title: Sequence of membrane-associated thyroid hormone binding protein from bovine live
 A;Reference number: A26829; MUID:87298601; PMID:3619939
 A;Accession: A26829
 A;Molecule type: mRNA
 A;Residues: 1-510 <YAM>
 A;Cross-references: UNIPROT:P05307; GB:M17596; NID:G163496; PIDN:AAA30690.1; PID:G163497
 A;Note: the authors translated the codon TTC for residue 175 as Leu
 C;Comment: This enzyme catalyzes the rearrangement of both intrachain and interchain dis
 C;Superfamily: protein disulfide-isomerase; thioredoxin homology
 C;Keywords: ascorbic acid; duplication; endoplasmic reticulum; homodimer; intramolecular
 F;1-19/Domain: signal sequence #status predicted <SIG>
 F;20-510/Product: protein disulfide-isomerase #status predicted <MAT>
 F;34-121/Domain: thioredoxin homology <TX1>
 F;377-462/Domain: thioredoxin homology <TX2>
 F;507-510/Region: endoplasmic reticulum retention signal
 F;55-58; 399-402/Disulfide bonds: redox-active #status predicted
 F;314-345/Disulfide bonds: #status predicted
 Query Match 65.1%; Score 41; DB 1; Length 510;
 Best Local Similarity 87.5%; Pred. No. 50;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 3 GSGDDDDL 10
 Db 479 GAGDDDDL 486
 |:|||||
 RESULT 14
 T46169
 CDC48-like protein - Arabidopsis thaliana
 N;Alternate names: protein T4D2.160
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
 C;Accession: T46169
 R;Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.; I
 submitted to the Protein Sequence Database, December 1999
 A;Reference number: Z23025
 A;Accession: T46169
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-815 <NYA>
 A;Cross-references: UNIPROT:Q9SCN8; EMBL:AL132958
 A;Experimental source: cultivar Columbia; BAC clone T4D2
 C;Genetics:
 A;Map position: 3
 A;Introns: 10/2; 65/3; 142/1; 397/3; 589/3
 A;Note: T4D2.160
 C;Superfamily: transitional endoplasmic reticulum ATPase; FtsH/SEC18/CDC48-type ATP-bind:
 Query Match 65.1%; Score 41; DB 2; Length 815;
 Best Local Similarity 77.8%; Pred. No. 82;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGDDDDL 10
 ||: |||||
 Db 805 GGAADDDL 813

RESULT 15

S66672

Phosphatidylinositol-specific phospholipase C - northern European squid
 C;Species: Loligo forbesi (northern European squid)
 C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
 C;Accession: S66672; S74310; S41169
 R;Carne, A.; McGregor, R.A.; Bhatia, J.; Sivaprasadarao, A.; Keen, J.N.; Davies, A.; Fin
 PEBS Lett. 372, 243-248, 1995
 A;Title: A beta-subclass phosphatidylinositol-specific phospholipase C from squid (Loligo
 forbesi)
 A;Reference number: S66672; MUID:96000215; PMID:7556677
 A;Accession: S66672
 A;Molecule type: mRNA
 A;Residues: 1-875 <CAR>
 A;Cross-references: UNIPROT:Q25285; EMBL:X76968; NID:9439518; PIDN:CAA54275.1; PID:94395
 A;Note: the authors translated the codon AAA for residue 86 as Asp and AAT for residue 1
 A;Accession: S74310
 A;Molecule type: protein
 A;Residues: 90-102;278-297;756-770 <CAN>
 C;Superfamily: Phosphatidylinositol-specific Phospholipase C; 1-phosphatidylinositol-4,5
 main Y homology

Query Match 65.1%; Score 41; DB 1; Length 875;

Best Local Similarity 77.8%; Pred.No. 88;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 GSGDDDDL 11
 ||: |||||
 Db 476 GAGDDELA 484

Search completed: May 27, 2005, 13:48:41

Job time : 3.92111 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:35:35 ; Search time 0.921109 Seconds
(without alignments)
972.511 Million cell updates/sec

Title: US-10-622-893A-1
Perfect score: 63
Sequence: 1 MGSGDDDDLLAL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	69.8	265	4	US-09-248-796A-20812
2	43	68.3	408	4	US-09-248-796A-19565
3	42	66.7	151	4	US-09-270-767-39203
4	42	66.7	151	4	US-09-270-767-54420
5	42	66.7	264	4	US-09-248-796A-27609
6	42	66.7	409	4	US-09-540-236-3384
7	42	66.7	543	4	US-09-248-796A-22504
8	42	66.7	1138	4	US-09-252-991A-25952
9	41	65.1	390	4	US-09-949-016-11598
10	41	65.1	445	4	US-09-252-991A-30715
11	41	65.1	491	1	US-07-872-673B-3
12	41	65.1	508	4	US-09-807-258-30
13	41	65.1	509	2	US-08-557-122A-29
14	41	65.1	509	3	US-09-262-666-29
15	41	65.1	510	2	US-08-557-122A-28
16	41	65.1	510	2	US-08-557-122A-36
17	41	65.1	510	3	US-09-262-666-28
18	41	65.1	510	3	US-09-262-666-36
19	41	65.1	675	4	US-09-252-991A-27987
20	41	65.1	3052	2	US-08-557-122A-26
21	41	65.1	3052	3	US-09-262-666-26
22	40	63.5	10	4	US-08-969-334-1
23	40	63.5	11	1	US-08-200-900A-37
24	40	63.5	11	4	US-08-794-042-37
25	40	63.5	23	3	US-09-217-228-3
26	40	63.5	106	4	US-09-902-540-15731
27	40	63.5	114	3	US-09-217-228-5

```

28 40 63.5 134 3 US-09-272-342B-5 Sequence 5, Appli
29 40 63.5 154 3 US-09-439-313-383 Sequence 383, App
30 40 63.5 154 3 US-09-352-616A-383 Sequence 383, App
31 40 63.5 154 4 US-09-636-215-383 Sequence 383, App
32 40 63.5 154 4 US-09-685-166A-383 Sequence 383, App
33 40 63.5 154 4 US-09-679-426-383 Sequence 383, App
34 40 63.5 154 4 US-09-759-143-383 Sequence 383, App
35 40 63.5 154 4 US-09-651-236-383 Sequence 383, App
36 40 63.5 161 4 US-09-636-215-846 Sequence 846, App
37 40 63.5 161 4 US-09-685-166A-846 Sequence 846, App
38 40 63.5 161 4 US-09-679-426-846 Sequence 846, App
39 40 63.5 161 4 US-09-759-143-846 Sequence 846, App
40 40 63.5 161 4 US-09-651-236-846 Sequence 846, App
41 40 63.5 177 3 US-09-272-342B-6 Sequence 6, Appli
42 40 63.5 290 4 US-09-248-796A-14651 Sequence 14651, A
43 40 63.5 291 3 US-09-328-869-2 Sequence 2, Appli
44 40 63.5 291 3 US-09-629-774A-2 Sequence 2, Appli
45 40 63.5 296 1 US-07-745-382-14 Sequence 14, Appli

```

ALIGNMENTS

```

RESULT 1
US-09-248-796A-20812
; Sequence 20812, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20812
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-248-796A-20812

```

```

Query Match 69.8%; Score 44; DB 4; Length 265;
Best Local Similarity 77.8%; Pred No. 16;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 2 GSGGDDDDL 10
Db 202 GGGGDDDDI 210

```

```

RESULT 2
US-09-248-796A-19565
; Sequence 19565, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19565
; LENGTH: 408
; TYPE: PRT

```

```
; ORGANISM: Candida albicans
US-09-248-796A-19565

Query Match      68.1%; Score 43; DB 4; Length 408;
Best Local Similarity 72.7%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GGGSGDDDDLAL 12
      || ||||| |
Db      386 GGGGDDDDNGL 396

RESULT 3
US-09-270-767-39203
; Sequence 39203, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39203
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-39203

Query Match      66.7%; Score 42; DB 4; Length 151;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GGGSGDDDD 9
      || ||||| |
Db      68 GGGGDDDD 75

RESULT 4
US-09-270-767-54420
; Sequence 54420, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 54420
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-54420

Query Match      66.7%; Score 42; DB 4; Length 151;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GGGSGDDDD 9
      || ||||| |
Db      68 GGGGDDDD 75

RESULT 5
US-09-248-796A-27609
; Sequence 27609, Application US/09248796A
```

```
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 27609
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-27609

Query Match      66.7%; Score 42; DB 4; Length 264;
Best Local Similarity 87.5%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GGGSGDDDD 9
      || ||||| |
Db      222 GGGGDDDD 229

RESULT 6
US-09-540-236-3384
; Sequence 3384, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3384
; LENGTH: 409
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-3384

Query Match      66.7%; Score 42; DB 4; Length 409;
Best Local Similarity 87.5%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GGGSGDDDD 9
      || ||||| |
Db      386 GGGGDDDD 393

RESULT 7
US-09-248-796A-22504
; Sequence 22504, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 22504
; LENGTH: 543
```

; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-22504

Query Match 66.7%; Score 42; DB 4; Length 543;
Best Local Similarity 87.5%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGGDDDD 9
Db 298 GGGGDDDD 305

RESULT 8

US-09-252-991A-25952
; Sequence 25952, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 25952

; LENGTH: 1138

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-25952

Query Match 66.7%; Score 42; DB 4; Length 1138;
Best Local Similarity 63.6%; Pred. No. 1.5e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GSGGDDDDDLAL 12
Db 303 GGQGDQDLAI 313

RESULT 9

US-09-949-016-11598

; Sequence 11598, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; PRIOR FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: Fast-SEQ for Windows Version 4.0

; SEQ ID NO 11598

; LENGTH: 390

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-11598

Query Match 65.1%; Score 41; DB 4; Length 390;
Best Local Similarity 87.5%; Pred. No. 69;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GSGDDDDL 10
Db 359 GAGDDDDL 366

RESULT 10

US-09-252-991A-30715

; Sequence 30715, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 30715

; LENGTH: 445

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30715

Query Match 65.1%; Score 41; DB 4; Length 445;
Best Local Similarity 58.3%; Pred. No. 79;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MGGSGDDDDLAL 12
Db 190 LAGSGADDDIAV 201

RESULT 11

US-07-872-673B-3

; Sequence 3, Application US/07872673B

; Patent No. 5578466

; GENERAL INFORMATION:

; APPLICANT: Toshiya HAYANO, Setsuko KATO, No. 5578466uhiro TAKAHASHI, and Masanori

; TITLE OF INVENTION: Co-expression System of Protein Disulfide Isomerase Gene an

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS & CUSHMAN

; STREET: 130 WATER STREET

; CITY: BOSTON

; STATE: MASSACHUSETTS

; COUNTRY: US

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Type 2DD, 3.50 inch, 720 KB

; COMPUTER: Apple Macintosh SE

; OPERATING SYSTEM: Apple DOS

; SOFTWARE: Microsoft Word Version 4.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/872,673B

; FILING DATE: 19920417

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: Japanese Patent Application No. 5578466. 114074/91 and 31160

; FILING DATE: 18-APR-1991 and 30-OCT-1991

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 491

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; DESCRIPTION: protein

US-07-872-673B-3

Query Match 65.1%; Score 41; DB 1; Length 491;

Best Local Similarity 87.5%; Pred. No. 88;					
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;					
QY	3	GSRDDDDL 10			
		:			
Db	460	GAGDDDDL 467			
RESULT 12					
US-09-807-258-30					
Sequence 30, Application US/09807258					
Patent No. 6670166					
GENERAL INFORMATION:					
APPLICANT: E. I. du Pont de Nemours and Company					
TITLE OF INVENTION: Arthropod Protein Disulfide Isomerases					
FILE REFERENCE: BB-1253 PCT					
CURRENT APPLICATION NUMBER: US/09/807,258					
CURRENT FILING DATE: 2001-06-11					
PRIOR APPLICATION NUMBER: 60/104,376					
PRIOR FILING DATE: 1998-10-15					
NUMBER OF SEQ ID NOS: 32					
SOFTWARE: Microsoft Office 97					
SEQ ID NO 30					
LENGTH: 508					
TYPE: PRT					
ORGANISM: Homo sapiens					
US-09-807-258-30					
Query Match 65.1%; Score 41; DB 4; Length 508;					
Best Local Similarity 87.5%; Pred. No. 91;					
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;					
QY	3	GSRDDDDL 10			
		:			
Db	477	GAGDDDDL 484			
RESULT 13					
US-08-557-122A-29					
Sequence 29, Application US/08557122A					
Patent No. 5879664					
GENERAL INFORMATION:					
APPLICANT: Hjort, Carsten Mailand					
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase					
NUMBER OF SEQUENCES: 38					
CORRESPONDENCE ADDRESS:					
ADDRESSEE: No. 5879664o No. 5879664disk of No. 5879664th America, Inc.					
STREET: 405 Lexington Avenue, 64th Floor					
CITY: New York					
STATE: New York					
COUNTRY: United States of America					
ZIP: 10174-6401					
COMPUTER READABLE FORM:					
MEDIUM TYPE: Floppy disk					
COMPUTER: IBM PC compatible					
OPERATING SYSTEM: PC-DOS/MS-DOS					
SOFTWARE: Patent In Release #1.0, Version #1.30					
CURRENT APPLICATION DATA:					
FILING DATE: 11-DEC-1995					
CLASSIFICATION: 435					
ATTORNEY/AGENT INFORMATION:					
NAME: Lambiris, Elias J.					
REGISTRATION NUMBER: 33,728					
REFERENCE/DOCKET NUMBER: 3980.204-US					
TELEPHONE: 212-867-0123					
TELEFAX: 212-878-9655					
INFORMATION FOR SEQ ID NO: 29:					
SEQUENCE CHARACTERISTICS:					
LENGTH: 509 amino acids					
TYPE: amino acid					
STRANDEDNESS: single					
TOPOLOGY: linear					
MOLECULE TYPE: peptide					
US-09-262-666-29					
Query Match 65.1%; Score 41; DB 3; Length 509;					
Best Local Similarity 87.5%; Pred. No. 91;					
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;					
QY	3	GSRDDDDL 10			
		:			
Db	478	GAGDDDDL 485			
RESULT 15					
US-08-557-122A-28					
Sequence 28, Application US/08557122A					
Patent No. 5879664					
GENERAL INFORMATION:					
APPLICANT: Hjort, Carsten Mailand					
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase					
NUMBER OF SEQUENCES: 38					
TYPE: amino acid					
STRANDEDNESS: single					

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: No. 58796640 No. 5879664disk of No. 5879664th America, Inc.
;; STREET: 405 Lexington Avenue, 64th Floor
;; CITY: New York
;; STATE: New York
;; COUNTRY: United States of America
;; ZIP: 10174-6401
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/557,122A
;; FILING DATE: 11-DEC-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Lambiris, Elias J.
;; REGISTRATION NUMBER: 33,728
;; REFERENCE/DOCKET NUMBER: 3980.204-US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212-867-0123
;; TELEFAX: 212-878-9655
;; INFORMATION FOR SEQ ID NO: 28:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 510 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-08-557-122A-28

Query Match 65.1%; Score 41; DB 2; Length 510;
Best Local Similarity 87.5%; Pred. No. 91;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GSGDDDDL 10
|:|||||
Db 479 GAGDDDDL 486

Search completed: May 27, 2005, 13:49:57
Job time : 1.92111 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:28:55 ; Search time 95.9723 Seconds
(without alignments)
2331.700 Million cell updates/sec

Title: US-10-622-893A-3
Perfect score: 2329
Sequence: 1 AVTKSSLLIVGAGTGTST.....MDFHDVKEWTVNQYRDISKL 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2329	100.0	438	2 P78573	P78573 aspergillus
2	1981.5	85.1	437	2 Q92204	Q92204 aspergillus
3	1242	53.3	445	2 O42629	O42629 aspergillus
4	848	36.4	426	2 O6B280	O6B280 debaryomyce
5	846.5	36.3	418	2 O6B2A7	O6B2A7 debaryomyce
6	660	28.3	437	2 Q92261	Q92261 penicillium
7	629	27.0	438	2 Q96074	Q96074 emericella
8	625	26.8	437	2 Q765A9	Q765A9 eupenicilli
9	600	25.8	437	2 Q765A8	Q765A8 coniochaeta
10	488.5	21.0	412	2 O43029	O43029 schizosacch
11	465	20.0	433	2 Q9UTW9	Q9UTW9 schizosacch
12	426.5	18.3	502	2 Q78638	Q78638 neurospora
13	386	16.6	433	2 Q6CL10	Q6CL10 kluyveromyc
14	376.5	16.2	486	2 Q6BN43	Q6BN43 debaryomyce
15	351	15.1	446	2 Q6BN42	Q6BN42 debaryomyce
16	346	14.9	435	2 O9P976	O9P976 cylindrocarr
17	192	8.2	390	1 MSOX_BACB0	P40859 bacillus sp
18	185	7.9	387	2 Q6ITC6	Q6ITC6 bacillus sp
19	183	7.9	371	2 Q737A8	Q737A8 bacillus ce
20	181.5	7.8	440	2 Q9V9B5	Q9V9B5 drosophila
21	181	7.8	384	1 MSOX_STRAW	Q827H4 streptomyc
22	181	7.8	388	1 MSOX_STRSB	P40854 streptomyc
23	177.5	7.6	390	1 SOX_HUMAN	Q9P029 homo sapien
24	177.5	7.6	390	2 Q6IAJ9	Q6IAJ9 homo sapien
25	175	7.5	371	2 Q63AM4	Q63AM4 bacillus ce
26	174	7.5	371	2 Q81PT2	Q81PT2 bacillus an
27	172.5	7.4	395	2 Q92NL2	Q92NL2 rhizobium m
28	172	7.4	367	2 Q982V8	Q982V8 rhizobium l
29	170.5	7.3	773	2 Q66IN9	Q66IN9 xenopus lae
30	170	7.3	371	2 Q8ICN0	Q8ICN0 bacillus ce
31	170	7.3	371	2 Q6H111	Q6H111 bacillus th

ALIGNMENTS

RESULT 1

```
P78573 ID P78573 PRELIMINARY; PRT; 438 AA.
AC P78573;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructosyl amine: oxygen oxidoreductase.
OS Aspergillus fumigatus (Sartorya fumigatus).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97284723; PubMed=9139700; DOI=10.1074/jbc.272.19.12505;
RA Takahashi M., Pischetsrieder M., Monnier V.M.;
RT "Molecular cloning and expression of anadoriase isoenzyme (fructosyl
aminoxidoreductase, EC 1.5.3) from Aspergillus fumigatus.";
RL J. Biol. Chem. 272:12505-12507(1997).
DR EMBL; U82830; AAC49711.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR006076; Fad oxred.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR000594; ThiF_domain.
DR Pfam; PF01266; DAO; 1.
SQ SEQUENCE 438 AA; 48931 MW; FBC8A3ESD89D02AE CRC64;
```

Query Match 100.0%; Score 2329; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.9e-174; Indels 0; Gaps 0;
Matches 437; Conservative 0; Mismatches 0;

```
Qy 1 AVTKSSLLIVGAGTGTSTALHLARRGYTNVTVDLPVPSPSAISAGNDVNVKVISSGQYS 60
Db 2 AVTKSSLLIVGAGTGTSTALHLARRGYTNVTVDLPVPSPSAISAGNDVNVKVISSGQYS 61
Qy 61 NNKDEIEVNEILAEAFNGWKNDFLPKPYHTDTGLLSACSQEGLDRLGVVRGEGDNL 120
Db 62 NNKDEIEVNEILAEAFNGWKNDFLPKPYHTDTGLLSACSQEGLDRLGVVRGEGDNL 121
Qy 121 VELTRPQRFKLAPEGVLQGFQWKGCFARSGAGWAHARNALVAAAREQRMGVKFTVG 180
Db 122 VELTRPQRFKLAPEGVLQGFQWKGCFARSGAGWAHARNALVAAAREQRMGVKFTVG 181
Qy 181 TPQGRVVTLIFENNNDVKAGVTGDKIWRARTFLCAGASAGQFLDFKNQLRPTATLWHI 240
Db 182 TPQGRVVTLIFENNNDVKAGVTGDKIWRARTFLCAGASAGQFLDFKNQLRPTATLWHI 241
Qy 241 ALKPEERALKYNIPIVNIERGFPPFEPDEERGEIKICDEHPGYTNMVQSDGTWMSIPPE 300
Db 242 ALKPEERALKYNIPIVNIERGFPPFEPDEERGEIKICDEHPGYTNMVQSDGTWMSIPPE 301
Qy 301 KTOIPKEAEFVRALLKETMPQLADRPFSFARICWCADTANREFLIDRHPQYHSLVLCGC 360
Db 302 KTOIPKEAEFVRALLKETMPQLADRPFSFARICWCADTANREFLIDRHPQYHSLVLCGC 361
```

```
32 170 7.3 399 2 Q95U69 Q95U69 drosophila
33 169.5 7.3 390 1 SOX_RABIT P79371 oryctolegus
34 166 7.1 388 1 MSOX_ARTST P40873 arthrobaete
35 166 7.1 435 2 Q8XTY8 Q8XTY8 ralstonia s
36 165 7.1 382 2 Q88GE9 Q88GE9 pseudomonas
37 162.5 7.0 390 2 Q8N6Z6 Q8N6Z6 homo sapien
38 160 6.9 390 1 SOX_MOUSE Q8D826 mus musculu
39 160 6.9 423 1 YGCN_ECOLI Q46904 escherichia
40 159.5 6.8 381 2 Q82M71 Q82M71 streptomyce
41 159.5 6.8 410 2 Q89M92 Q89M92 bradyrhizob
42 155 6.7 331 2 Q6R5L2 Q6R5L2 bacillus sp
43 155 6.7 375 2 Q8EMP0 Q8EMP0 oceanobacil
44 155 6.7 387 1 MSOX_BACSN P23342 bacillus sp
45 153 6.6 433 2 Q8FEI5 Q8FEI5 escherichia
```


NCBI_TaxID=284592;
[1]
SEQUENCE FROM N.A.
RC STRAIN=CBS767;
Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Tallia E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cattolico L., Confanier F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jaumiaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RA Genoscope;
RA SubMITTED (JUL-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL; CR382133; CAG84411.1; -;
DR InterPro; IPR006076; FAD_OXRD.
DR InterPro; IPR000205; NAD_BS.
DR Pfam; PF01266; DAO; 1.
SQ SEQUENCE 426 AA; 48297 MW; 7A5A382DF172293 CRC64;

Query Match 36.4%; Score 848; DB 2; Length 426;
Best Local Similarity 41.2%; Pred. No. 4.4e-58;
Matches 181; Conservative 71; Mismatches 153; Indels 34; Gaps 10;

Qy 4 KSSLLIVGAGTGTSTALHLARRGYNTVTLDPVPVPSAISAGNDVNVKVISSGQ----Y 59
Db 3 KPGKILLIAGTGTSTALHLARRQGEKDVILDPVAVPSPFAGNDVNVKIIQTTSDDDFY 62

Qy 60 SNNKDIENVNIIAEAFNGKNDPLFKPYHYDHTGLMSACSQEGLDRLGVVR--PGE 117
Db 63 SK-----LALALEWREDNVFNKAFATGTGIIVTAATGKQESIDRYEYLLGRK 112

Qy 118 PNLVLTREPOFKLAP--EGV--LQDPPGKGYFARSAGWAHARNALVAAAREAGRM 173
Db 113 DKVVLNSVEDYKYPNKEGLKSYNKKYGYOYKNGCWAFARLALENCVECKL 172

Qy 174 GVKFTGTGQVRVTLIF--ENNDVKGAVTDGKIMWRAETFLCAGASQGLDFKNQLRP 232
Db 173 GAKFVIDSAB---ELLFSEDGACVHTSNGNIETADRTIICAGANSFKFLNFEQQLA 228

Qy 233 TANTLHVIALKPERALYKNIPVIFNIERGFPEDEERGEIKIDEHPGYTNMVQSADG 292
Db 229 KYCTLGHILKLTDEAALLKGMVPLNDGCFVFPED--LNNEIKFCNCFPGYINVED-- 285

Qy 293 TMSIPEKTOIPKEAETRVALLKETMPQADRPSPFARIWCACADTANREFLIDRHPQY 352
Db 286 ---SVPSFKDSIPKEADQWAFRLQVFPFARERESLARIWCCTDTDPDRFLICEHPGH 342

Qy 353 HSLVLCGASGRGFKYLPISGNLIVDM---EGKVPQKIHLEIKWNPDIANRNWRDTLG 409
Db 343 KNLVLGTGDSGQGYKMPNVGKVISQVALKGENSLDKKELWLRWPDWKKRDLKDLQ 402

Qy 410 RFGGPRVMDFDHVKWNTN 428
Db 403 RYGGSEVNDLKNVQWSN 421

RESULT 5
Q6B2A7 PRELIMINARY; PRT; 418 AA.
ID Q6B2A7
AC Q6B2A7;

RESULT 5
Q6BZA7
ID Q6B
AC Q6B

QY 281 PGYTNMQS---ADGTMMSIPP---EKTQIPKEATRVALLKETMPQLADRPSPFARIC 334
 Db 285 YEYVCNVETDYSKVSVPISPHSGPSKSLPKYAIQWRRFLDTFLPDADRSLLNTKMC 344
 QY 335 WCADTANREFLIDRHPQYHSLVLCGASGRGFKYLPISIGNLIVDAMRGKVPQKIHILIKW 394
 Db 345 WISDTEANFLIDKVPQFDNVFVANGDSGHAFKFLPNIGRYIAQRIIGDLISEWKDAWRW 404
 QY 395 NPDIANR-NWR 405
 Db 405 REDDKSELKWR 416

RESULT 12
 Q7S638
 ID Q7S638 PRELIMINARY; PRT; 502 AA.
 AC Q7S638;
 DT 01-MAR-2004 (Tremblrel. 26, Created)
 DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
 DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)
 DE Hypothetical protein (Related to fructosyl amino acid oxidase).
 GN Name:NCU04771.1; Synonyms=B15810.150;
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OR74A;
 RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D., Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B., Elkins T., Engels S., Wang S., Nielsen C.B., Butler J., Endrizzi M., Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M., Selikrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U., Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D., Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S., Kamal M., Kamyselis M., Mauceli E., Bielek C., Rudd S., Frisman D., Kryzofova S., Rasmussen C., Merzenberg R.B., Perkins D.D., Kroken S., Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A., DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R., Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R., Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M., Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
 RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
 RL Nature 0:0-0(2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Partmann B., Holland R., Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.
 CC EMBL; AABX01000331; EAA30984.1; -;
 DR EMBL; BX897677; CAE85581.1; -;
 DR InterPro; IPR006076; Pad_oxred.
 DR Pfam; PF01266; DAO; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 502 AA; 52930 MW; 39748F9B553D4FDE CRC64;

Query Match 18.3%; Score 426.5; DB 25; Length 502;
 Best Local Similarity 29.2%; Pred. No. 7.3e-25;
 Matches 136; Conservative 56; Mismatches 169; Indels 109; Gaps 21;
 QY 6 SSLIIVGAGTWGTSALHLARR-GYTNVTVL-----DPYVPVPSAISAGNDVNKVI 54
 Db 30 SSLIIVGAGTWGTSALHLARR-GYTNVTVL-----DPYVPVPSAISAGNDVNKVI 54
 Db 30 SSLIIVGAGTWGTSALHLARR-GYTNVTVL-----DPYVPVPSAISAGNDVNKVI 54

QY 55 SSGOYNNKDEIEVNEILAEAEFNGWK-----NDPLPKPYVYHDTGLLMSACSQSG-LDRL 108
 Db 88 RA-DYPD-----AAYAALAAEAQLQWRQOHTDPLDSEGRYSSEGLLWA---DGPAPPP 138
 QY 109 GVRVRPG-----EDNLV-----ELTRPEQFRKLAPEGV 137
 Db 139 GTPVAPGTSVIDKSLKTGM DYARPSWANVLSSASSDPELAARIKECPNTDAIKALGTG- 197
 QY 138 LQGDFFPGWKGYFARGAGWAHARNALVAAAREAQRMG-VKPVGTGTPGVRVTLIFENNDV 196
 Db 198 --GSSGSW-GYI-NGLSGWANAGASMAWLYKRVRAEGRINFVA---GEVTNLEVSNGTV 249
 QY 197 KGAVTGDGKIWRARTFLCAGASAGQFLDFKNQLRPTAATLTVHIALKEEERALKYNTPI 256
 Db 250 TGAISDGRVLSADLVNVSAGAWTGRVLDTGTQAIATGQVLGYIDLTPESBAQLAHPVI 309
 QY 257 FNIERGFFFPDEBERGEIKICDEHPGYTNMVQSADGTMMSIP-----FEKT----- 302
 Db 310 LNLSTGLFVIP-PRNGVLKVARHAYGYLN-----PTTLSVPPPLASFVTTTAAAVSLPLT 362
 QY 303 -----QIPKEATRVRLAKE--TWPQLADRPSPFARICWCADTANREFLIDRHPQYH 353
 Db 363 TLTDPTLQIPTGADDLRRALHEMVPLPSLRDRPFSKTRICWYSDTPTADFIVDYHPK 422
 QY 354 SILVLCGASGRGKYLPSIGNLIVDAMEGKVPQKIHILIKWNPDIANRNR 405
 Db 423 GLFVATGDSGHAFKFLPVIAGEKIADVIAGQCP-----PEFVGKWNMR 464

RESULT 13
 Q6CL10
 ID Q6CL10 PRELIMINARY; PRT; 433 AA.
 AC Q6CL10;
 DT 25-OCT-2004 (Tremblrel. 28, Created)
 DT 25-OCT-2004 (Tremblrel. 28, Last sequence update)
 DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)
 DE Similar to sp|Q9UTM9 Schizosaccharomyces pombe Putative fructosyl amino acid oxidase.
 DE ORFNames=KLIA0F02860g;
 OS Kluyveromyces lactis NRRL Y-1140.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetaceae; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_TaxID=284590;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL Y-1140;
 RG Genolevures;
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neugeglise C., Talla E., Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C., Boisrame A., Boyer J., Cattolico L., Confaniolieri F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Karrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller R., Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O., Pellens S., Potier S., Richard G.F., Straub M.L., Suleau A., Swenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zenou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J., Wincker P., Souciet J.L.;
 RT "Genome evolution in yeasts.";
 RL Nature 430:35-44(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL Y-1140;
 RA Genoscope;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; CR382126; CAG97917.1; -;
 DR InterPro; IPR006076; Pad_oxred.
 DR InterPro; IPR000205; NAD_BS.
 DR Pfam; PF01266; DAO; 1.
 SQ SEQUENCE 433 AA; 48905 MW; 1091A3C6C3F3DA08 CRC64;


```
RA Boissrame A., Boyer J., Cattolico L., Confaniolieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennens D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Boiotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RC SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: CR382138; CAG88680.1; -.
DR InterPro: IPR006076; Rad oxd.
DR InterPro: IPR00205; NAD_BS.
DR Pfam: PF01266; DAO; 1.
SQ SEQUENCE 446 AA; 50897 MW; 6C89A7DBBAB75C6B CRC64;

Query Match      15.1%; Score 351; DB 2; Length 446;
Best Local Similarity 26.8%; Pred. No. 5.4e-19;
Matches 126; Conservative 74; Mismatches 186; Indels 84; Gaps 22;

QY      5 SSSLIIVGAGTWTGTSTALHLARRGYTNVT--VLDPY--PVPSAISAGNDVKNVISSGQYS 60
DB      4 TKDIIIVGAGIFGLSTALAEER-YANRTHIVDRFEPFVPGSSV--DTTRCIR-----P 56

QY      61 NKKDEIEVNEILAEAFNGWKNDFLPKPYHDTGL-----LMSACSOEGL 105
DB      57 DYKDETYCR--LAHESMNLKADPMISKLFHQOQMTFVYDGNHDTWEDIPLSSKSSAE-- 112

QY      106 DRLGVRVRPGEDPNLVELTRPEQ-FRKL---APEGLVQGDPPG---W-KGYFARSAGWA 157
DB      113 -----KVNREEEGKMDYYDSAEDEVFSIHNRSPNSVETTELGGPKKWKNGYTNRVN-GFI 166

QY      158 HARNAL---VAAREAQRMGVKFTGTGPGRVVTVLI FENNDVKGAVTGDGKIWRARTFL 214
DB      167 DARASIKVYYERARNFPNVNFRF---QVEKIDYFKNTKKVGLKNGDKISADLVIV 222

QY      215 CAGASAGQFLDFKNQLRPTAWTLVHIALKPEERALYKNIPIVFNIERGF-FFEPDEERGE 273
DB      223 AAGAWSCKLVNLDNISKSSAIEVVWFKVTPSMEKEWQMSITTNLTGINIFPP--YNGE 280

QY      274 IKICDEHPGYTNMQV-----SADGTMSIPEKTOIPKEAETRVRLIKET 319
DB      281 IKVLRRSAGYKNTVAIESPSPFSRVKQEISVPTIIDSP--KOWIPEAEIEINGNLKEI 338

QY      320 MPOLADRPFGPARICWCADTANREFLIDRHPQHSVLVGGCAGRGFKYLPISIGNLIYDA 379
DB      339 MPSLCELPFRTKLCWLTQTTSNFIIDYPEAENVLVATGSSAHAWKFPVIGDKVVDV 398

QY      380 MEGKVPQKIHILIKWNPDI AANRNWRD-----TLGRFGGPNRVMDFDHVK 425
DB      399 IEGRLTELEL--IRKW-----SWKEKLTSTVENENAPRMKGDPKEMKE 438
```

Search completed: May 27, 2005, 13:47:27
Job time : 98.9723 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:35:35 ; Search time 33.5437 Seconds
(without alignments)
972.511 Million cell updates/sec

Title: US-10-622-893A-3

Perfect score: 2329

Sequence: 1 AVTKSSLLIVGAGTGTST.....MDFHDKWENVQYRDISKL 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1981.5	85.1	437	2	US-09-031-059-1
2	1981.5	85.1	437	2	US-09-031-059-3
3	160	6.9	433	4	US-09-711-164-412
4	155	6.7	387	3	US-09-457-302-1
5	148.5	6.4	866	4	US-09-949-016-10219
6	146.5	6.3	448	4	US-09-252-991A-21043
7	135.5	5.8	378	3	US-09-199-637A-367
8	135.5	5.8	419	4	US-09-252-991A-21423
9	130.5	5.6	484	4	US-09-252-991A-24053
10	125	5.4	652	4	US-09-252-991A-30872
11	123.5	5.3	476	4	US-09-489-039A-12890
12	123.5	5.3	486	4	US-09-684-405-2
13	116.5	5.0	431	1	US-08-391-339-18
14	116.5	5.0	431	1	US-08-484-274A-18
15	115	4.9	505	4	US-09-902-540-9986
16	114.5	4.9	474	4	US-09-543-681A-6873
17	112.5	4.8	626	4	US-09-331-568A-27
18	110.5	4.7	375	4	US-09-328-352-6891
19	110.5	4.7	635	4	US-09-543-681A-7499
20	108	4.6	380	4	US-09-489-039A-7473
21	108	4.6	430	1	US-08-391-339-5
22	108	4.6	430	1	US-08-484-274A-5
23	106.5	4.6	486	4	US-09-543-681A-7036
24	106	4.6	448	4	US-09-489-039A-7529
25	104.5	4.5	496	3	US-09-147-009-8
26	104.5	4.5	649	4	US-09-252-991A-18996
27	101.5	4.4	438	4	US-09-252-991A-26523

28	101.5	4.4	477	4	US-09-252-991A-27656	Sequence 27656, A
29	100.5	4.3	433	4	US-09-684-405-12	Sequence 12, Appl
30	100.5	4.3	443	4	US-09-602-787A-80	Sequence 80, Appl
31	100.5	4.3	676	4	US-09-543-681A-5943	Sequence 5943, Ap
32	98	4.2	421	4	US-09-602-787A-246	Sequence 246, App
33	97.5	4.2	524	3	US-08-557-210A-3	Sequence 3, Appli
34	97	4.2	401	4	US-09-252-991A-17272	Sequence 17272, A
35	97	4.2	539	3	US-08-557-210A-5	Sequence 5, Appli
36	97	4.2	556	4	US-09-543-681A-5835	Sequence 5835, Ap
37	96.5	4.1	568	4	US-09-266-965-135	Sequence 135, App
38	96.5	4.1	765	4	US-09-252-991A-24791	Sequence 24791, A
39	96	4.1	359	4	US-09-302-540-14124	Sequence 14124, A
40	96	4.1	518	4	US-09-252-991A-18753	Sequence 18753, A
41	95.5	4.1	626	4	US-09-302-540-10385	Sequence 10385, A
42	95.5	4.1	642	4	US-09-328-352-5199	Sequence 5199, Ap
43	95.5	4.1	945	4	US-09-489-039A-8135	Sequence 8135, Ap
44	95	4.1	1274	4	US-09-252-991A-20386	Sequence 20386, A
45	94.5	4.1	558	4	US-09-252-991A-16908	Sequence 16908, A

ALIGNMENTS

RESULT 1
US-09-031-059-1
; Sequence 1, Application US/09031059
; Patent No. 5948659
; GENERAL INFORMATION:
; APPLICANT: KATO, NOBUO
; APPLICANT: SAKAI, YASUYOSHI
; APPLICANT: TANI, YOSHIKI
; APPLICANT: FUKUYA, HIROSHI
; TITLE OF INVENTION: RECOMBINANT FRUCTOSYL AMINO ACID
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,059
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/899,172
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, MARC S
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 0020-4253P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-031-059-1

Query Match 85.1%; Score 1981.5; DB 2; Length 437;
Best Local Similarity 82.6%; Pred. No. 1.4e-200;
Matches 360; Conservative 43; Mismatches 32; Indels 1; Gaps 1;

```
QY 2 VTKSSLLIVGAGTGTSTALHLARGYTNVTVDLPYPVPSAISAGNDVKNKVISSQYSN 61
Db 3 VTKSSLLIAGTGTSTALHLARGYTNVTVDLPYPVPSAISAGNDVKNKVISSQYSS 62
QY 62 NKDEIEVNEILAEAFNGWKNDPLFKPYHYDHTGLLMSACSOEGLDRLGVRVRPGEDEPNLV 121
Db 63 KKDEVEVNEIIAEAFNGWKNDPIFKPYHYDHTGVVMSATTQEGLERLGVRVRPEDEPDVA 122
QY 122 ELTRPEQFKLAPGVLQDFFGKGYFARSAGAGAHARNALVAAAREAQRMGVKFTVTGT 181
Db 123 ELTRPEQFKLAP-GVLEKGNFPGWRGYHRSNAGWAHARNALVAAAREAQRLGVRVAVGS 181
QY 182 PQGRVVTLLIFENNVDKAVTGDGKIWRABETFLCAGASAGQFLDFKQQLRPTAHTLVHIA 241
Db 182 PQGRVVTLLIFENNVDKAVTADGKIWRABETFLCAGAAAGQFLDFKQQLRPTAHTLVHIQ 241
QY 242 LKPEERALKYNIPIVNIERGFPEDEERGEIKICDEHPGYTNMVQSDGTMMSIPFEK 301
Db 242 LKPEERAQYKMPVFNIEKGFPEDEERGEIKICDEHPGYTNMTTGADGRVRSIPFEK 301
QY 302 TQIPKEATRVKALLKETWPLQADRPSPARICWCADTANREFLIDRHPQYHSLVLGCGA 361
Db 302 TQVPREAENVRKLLSETWPLQADRPSPARICWCADTANREFLIDRHPQYHSLVLGCGA 361
QY 362 SGRGFKYLPISIGNLIVDAMEGKVPQKIHELKWNPDIAANRNWRDNLGRFGGPNRVMDFH 421
Db 362 SGRGFKYLPISIGSIADAMEDKTPAKIHKLRWSPAIARNWGDRLGRFGGPNRVMDFN 421
QY 422 DVKEWTNVQYRDISKL 437
Db 422 EVKEWTNVQYRDISKL 437
```

RESULT 2

```
US-09-031-059-3
; Sequence 3, Application US/09031059
; Patent No. 5948659
; GENERAL INFORMATION:
; APPLICANT: KATO, NOBUO
; APPLICANT: SAKAI, YASUYOSHI
; APPLICANT: TANI, YOSHIKI
; APPLICANT: FUKUYA, HIROSHI
; TITLE OF INVENTION: RECOMBINANT FRUCTOSYL AMINO ACID
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,059
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/899,172
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, MARC S
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 0020-4253P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
```

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-031-059-3

Query Match 85.1%; Score 1981.5; DB 2; Length 437;
Best Local Similarity 82.6%; Pred. No. 1.4e-200;
Matches 360; Conservative 43; Mismatches 32; Indels 1; Gaps 1;

QY 2 VTKSSLLIVGAGTGTSTALHLARGYTNVTVDLPYPVPSAISAGNDVKNKVISSQYSN 61
Db 3 VTKSSLLIAGTGTSTALHLARGYTNVTVDLPYPVPSAISAGNDVKNKVISSQYSS 62
QY 62 NKDEIEVNEILAEAFNGWKNDPLFKPYHYDHTGLLMSACSOEGLDRLGVRVRPGEDEPNLV 121
Db 63 KKDEVEVNEIIAEAFNGWKNDPIFKPYHYDHTGVVMSATTQEGLERLGVRVRPEDEPDVA 122
QY 122 ELTRPEQFKLAPGVLQDFFGKGYFARSAGAGWAHARNALVAAAREAQRMGVKFTVTGT 181
Db 123 ELTRPEQFKLAP-GVLEKGNFPGWRGYHRSNAGWAHARNALVAAAREAQRLGVRVAVGS 181
QY 182 PQGRVVTLLIFENNVDKAVTGDGKIWRABETFLCAGASAGQFLDFKQQLRPTAHTLVHIA 241
Db 182 PQGRVVTLLIFENNVDKAVTADGKIWRABETFLCAGAAAGQFLDFKQQLRPTAHTLVHIQ 241
QY 242 LKPEERALKYNIPIVNIERGFPEDEERGEIKICDEHPGYTNMVQSDGTMMSIPFEK 301
Db 242 LKPEERAQYKMPVFNIEKGFPEDEERGEIKICDEHPGYTNMTTGADGRVRSIPFEK 301
QY 302 TQIPKEATRVKALLKETWPLQADRPSPARICWCADTANREFLIDRHPQYHSLVLGCGA 361
Db 302 TQVPREAENVRKLLSETWPLQADRPSPARICWCADTANREFLIDRHPQYHSLVLGCGA 361
QY 362 SGRGFKYLPISIGNLIVDAMEGKVPQKIHELKWNPDIAANRNWRDNLGRFGGPNRVMDFH 421
Db 362 SGRGFKYLPISIGSIADAMEDKTPAKIHKLRWSPAIARNWGDRLGRFGGPNRVMDFN 421
QY 422 DVKEWTNVQYRDISKL 437
Db 422 EVKEWTNVQYRDISKL 437
```

RESULT 3

```
US-09-711-164-412
; Sequence 412, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 412
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-412
```

```
Query Match 6.9%; Score 160; DB 4; Length 433;
Best Local Similarity 23.3%; Pred. No. 6.6e-08;
Matches 106; Conservative 49; Mismatches 146; Indels 154; Gaps 24;

QY 8 LLIVGAGTGTSTALHLARGYTNVTVDLPYPVPSAISAGNDVKNKVISSQYNNKDEIE 67
Db 17 IIIIGAGIAGTACALCARAG-LSVLLLEAEIPG-----SKNLSGRL----- 59
QY 68 VNEILAEAFNGWKNDPLFKPYHYDHTGLLMSACSOEGLDRL---GVRVRPGEDEPNLV 124
```

Db	60	YTHALAE								LLPQFHLTAPLERRITHESSLSLTPDGVTT	-----	96
Qy	125	RPEOFRKLAPBG	---	VLOQDFPGWKGYFARSSGAGWAHARNALVAAAREAQRMGVKFTVG	180							
Db	97	-----	FSSLQPGGESWSVLARFPDW	-----	---					---	---	130
Qy	181	TPQGRVVTLIPENNDVKAVTGDGKIWAERTFLCAGASAGQFLDPKNQLRPTAWTLVHI	240									
Db	131	PGATVDALYEENGRVCGVICGD	-DILRARYVVLAE	-----	---					NSVLAERHGLV	---	177
Qy	241	ALKPEERALYKNIPIVFN	-----	IERGFFPDPDEERGEI	---	KICDEHPG	---	YTNMVQ	288			
Db	178	TRPAGEAMALGIKEVLSLET	SAIEERPHLENNEGAALLFSGRICDDLP	CGNAFLYTNQOT	236							
Qy	289	SADG	-----	TWMSIPFEKTQIPKEATRVALLKET	-----	MPQ	---	LADRPF	329			
Db	237	LSLGI VCP	LSLTQSRVPASELLTRFKAHPAVRPLIKNTESLEYGAHLVPEGLHSPVQ	296								
Qy	330	PARTCW	-----	CADTANREFLIDRHPQYSLVLCGASGRGFKY	---	LPSIGNLI	376					
Db	297	YAGNGWLLVGDALRSCVNT	-----	GISVRGMDMALTGAQAAQT	336							
Qy	377	VDAMEGKVPQKIHILIKWNPDIANRWRTDILGRF	411									
Db	337	ISACOHREPOLNLFPL	--YHNHVERSLW-DVLORY	368								

```

RESULT 4
US-09-457-302-1
; Sequence 1, Application US/09457302
; Patent No. 6228626
; GENERAL INFORMATION:
; APPLICANT: ICHIKAWA, Toshio
; APPLICANT: KOYAMA, Yasuji
; TITLE OF INVENTION: SARCOSINE OXIDASE AND PROCESS FOR PRODUCING THE SAME
; FILE REFERENCE: 081356/0139
; CURRENT APPLICATION NUMBER: US/09/457,302
; CURRENT FILING DATE: 1999-12-09
; EARLIER APPLICATION NUMBER: JP 10-354482
; EARLIER FILING DATE: 1998-12-14
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-457-302-1

```

```

Qy 278 DEHPCYTNMVSADGTMWSIPFEKTQTPKEAETRVALLKKTMPOLADRPFSFARICWCA 337
Db 268 ----GYHSYGQQIDPDTINREFGAY---PEDEANLRKFEQYMPG-ANGELKKGAVCMYT 319
Qy 338 DTANREFLIDRHPQVHSLVLCGGASGRGFKYLPISIGNLIVD-AMEGKVPQKI 388
Db 320 KTPDEHFVIDLHPKYSVNAIAGFSGHGFKFSSVVGETLAQLATTGKTEHDI 371

RESULT 5
US-09-949-016-10219
; Sequence 10219, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USSS THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10219
; LENGTH: 866
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10219

```

RESULT 6
US-09-252-991A-21043

```
; Sequence 21043, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21043
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21043

Query Match          6.3%; Score 146.5; DB 4; Length 448;
Best Local Similarity 23.7%; Pred. No. 1.9e-06;
Matches 98; Conservative 59; Mismatches 164; Indels 93; Gaps 23;

QY 6 SSSLIVGAGTGTSTALHLARRGYTNVTVDLPYPVPSAISAGNDVNKVISSGQYNNKDE 65
Db 51 TDVCVIGAGTGLSTALFLENGF-KVTVLETAKVGFAS-GRNGGQIVNS--YSRDIDV 106
QY 66 IEVN-----EILAEAFNGWKNDPLFKPYHYDTGL-----LMSACSQEGLDRLGVR 111
Db 107 IERTVGKREAQLIGENAFEGGR---IIRERVARYIGCOLKGGVFPAFTEKQMDHLRAQ 163
QY 112 VRGE--DPNLVELTRPEQFRKLAPSGVLQDPPGKGYFARSGAGWAAHARNAVAAARE 169
Db 164 KQWERYGHNQLEIMDAKRIR-----EVATDNYIG--GMLDSG-GHIHPLNALGEAA 216
QY 170 AQRMGVKFTGTGQGRVVTLLIFENNNDVKGAVTGDGKIWAERTFLCAGASAGOLFDFKNQ 229
Db 217 VESLGGRIYEQSPATRI-----ERGASPVVHTPGKV---KAKFIVVAGNA--YLN--- 262
QY 230 LRPTATWLVIHIAKPERALYKNIP-----VI-----FNIERGFFPFDEBERGEIKICDEHP 281
Db 263 -----GLVPELAA--KSMPCGTQVITTEPLNEELAHSLLPQD-----YCVEDC 303
QY 282 GYTNMVQSGDGTWMSI-----PPEKTOIPKEAETRVALLKETMPOLADRPFSFARICWCA 337
Db 304 NYLLDYRLSGDKRLIYGGVVIYGARDPADIEAIRPKMLKTPPQDKDKIDFA---W--- 358
QY 338 DTANRFLIDRHPQ-----YHSLVLGCGASGRGFKYLPSTIGNLIVDAMEGK 383
Db 359 -TCNFLTLSRLPQVGRIGNDIYYSQ-----GCSGHGVYTHVAGKVLAEALRQ 407

RESULT 7
US-09-199-637A-367
; Sequence 367, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsengalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
```

```
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 367
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-367

Query Match          5.8%; Score 135.5; DB 3; Length 378;
Best Local Similarity 24.3%; Pred. No. 2.1e-05;
Matches 102; Conservative 45; Mismatches 165; Indels 107; Gaps 19;

QY 5 SSSLIVGAGTGTSTALHLARRGYTNVTVDLPYPVPSAISAGNDVNKVISSGQYNNK 64
Db 16 SRDVVVGAGVIGLLTARELALAGRLVTLVERGESGREASWAGGGIVSPLYPWRYS 72
QY 65 EIEVNEILAEAFNGWKNDPLFKPYHYDTGLLMSACSQEGLDRLGVRVPRGDPNL 120
Db 73 -----AVTALAHWSQD--FYP-----ALGQRLDDETGL-----DPEVHTVG 106
QY 121 ---VEL-----TRPEQFRKLAPSGVLQDPPGKGYFARS--GAGWAHARNA 162
Db 107 LWLDDDDQTEALQWANHTRP---LKEVPIEBAVAAVFLGAGFORAVYMSGVANRNP 163
QY 163 LVAAAREAQRMGVKFTGTGQGRVVTLLIFENNNDVKGAVTGDGKIWAERTFLCAGASAG 222
Db 164 RLARSILRASLQFANLELHEQTEVRGLRDGVRVGVATSRGEI-RGDKVLLAAGAWSGE 222
QY 223 FLDFKNQLRPTATWLVIHIAKPERALYK---NIPVIFNIEGFFPFDEBERGEIKICD 278
Db 223 L-----LKPLGLLELPVVPVK-QOMILYKCAADFLPRMV-LAKGRYAIPRD-GHILIGS 273
QY 279 --EHPGYTNMVQSGDGTWMSIPEKTOIPKEAETRVALLKETMPOLAD-----R 326
Db 274 TIEHSG-----FDKTPDTEALES-LRASAELLPELADMQPVAAHWAGLR 316
QY 327 PFSFARICWADTANREFLIDRHPQYHSLVLGCGASGRGFKYLPSTIGNLIVDAMEGKVP 385
Db 317 PGSPEGIPY-----IGVPVGFGLWLTNGHYRNLGLVLAPASCRLLADLMSGREP 365

RESULT 8
US-09-252-991A-21423
; Sequence 21423, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21423
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21423

Query Match          5.8%; Score 135.5; DB 4; Length 419;
Best Local Similarity 24.3%; Pred. No. 2.4e-05;
Matches 102; Conservative 45; Mismatches 165; Indels 107; Gaps 19;

QY 5 SSSLIVGAGTGTSTALHLARRGYTNVTVDLPYPVPSAISAGNDVNKVISSGQYNNK 64
Db 57 SRDVVVGAGVIGLLTARELALAGRLVTLVERGESGREASWAGGGIVSPLYPWRYS 113
QY 65 EIEVNEILAEAFNGWKNDPLFKPYHYDTGLLMSACSQEGLDRLGVRVPRGDPNL 120
```

Db 114 -----AVTALAHSQD--FYP-----ALGRLDDETGL-----DPEVTVG 147
Qy 121 ----VEL-----TRPEQFRKLAPGVQLQDPPGKMGFYARS--GAGWAHARNA 162
Db 148 LYLWLDLDDQTEALQWARKHTRP---LKEVPTIEEYAAVPLGAGFORAVYMSGVANVRNP 204
Qy 163 LVAAREARQMGVXFVTGTQGRVVTLIPENNVDKGVATGDKIWRABRTFLCAGASAGQ 222
Db 205 RLARSRLASLQQFANLHEQTEVRGWLDRDGRVLGVATSRGEI--RGDKVLLAAGAWSGE 263
Qy 223 FLDFKNQLRPATWTLVHIALKPEERALKYK-----NIPVFNIERGFFPEPEERGEIKICD 278
Db 264 L-----LKLGLLELPVVPVK--QWILYKCAADFLPMV--LAKRYAIPRRD--GHILIGS 314
Qy 279 --BHPGVTNMQSADGTMSTIPPEKTOIPKEATRVALLKETMPQLAD-----R 326
Db 315 TLEHSG-----FDKTPIDEALS--LRASAAELLPEDADMQPVVAHWAGLR 357
Qy 327 PFSFARICWCAADTANRFLIDRHPOYHSLVLCGASGRGFKYLPISGNLIVDAMEGKVP 385
Db 358 PGSPEGIPY-----IGPVFGDGLMLNTGHRNGLVLAPASCRLLADLMSGREP 406

RESULT 9
US-09-252-991A-24053
; Sequence 24053, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24053
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24053

Query Match 5.6%; Score 130.5; DB 4; Length 484;
Best Local Similarity 22.2%; Pred. No. 0.0001;
Matches 101; Conservative 70; Mismatches 183; Indels 101; Gaps 23;

Qy 4 KSSLLIVGAGTGTSTALHLAGRGYTNVTVDLPYPVPSAIS-----AGNDVYNK- 52
Db 81 EQADVCVGGFGSVNTALEAQRGP--SVLLEAHRIGWASGRNGGQLIRGVGHVDEQF 139
Qy 53 --VSSSQYNNKDEIEVNEILAEAFNGKNDPLFKPYHYDGTLLMSACSQEGDLRLGV 110
Db 140 FPVIGA-----DGVKALKMLGLEAVEIVRR--RVEQYAIIDCDLRWGYCD-----L 182
Qy 111 RVRRPG-----EDPNLVELTRPEQFRKLAP-----EGVLQGDFFQWKGYFARSAGAGWAHAR 160
Db 183 ANKPGDYQGFREDMEELQALGYRHEMLVPAAEHRSVVGSD--RYVGLVDMGSGHLHPL 240
Qy 161 NALVAAAREARQMGVXFVTGTQGRVVTLIPENNVDKGVATGDKIWRABRTFLCAGASA 220
Db 241 NLVLGEAAAQSLGVLRFERSP-----VTRIDYGAEVQ--VHTATGKV--RAKT--LVLGNA 292
Qy 221 GQFLDFKNQLRP-----TAWTLVHIALKP--ERALKYKNIPVFNIERGFFPEPEDEERG 272
Db 293 -----YNDNLNPLGGKVLFPAGSYV--IATEFLDQLARQLLP-----Q 329
Qy 273 EIKICDHPGVTNMQSADGTM--SIPPEKTOIPKEATRVALLKETMPQLADRPFSF 330
Db 330 NMAVCDQORVALDYRLSADNRLLPGGACHYSGRDPDSIAAYMRPKMLEVFPQLANVRIDY 389

Qy 331 ARICWCAADTANRFLIDRHPOYHSLVLCGASGRGFKYLPISGNLIVDAMEG----- 382
Db 390 QWGMIGIGANRLPQIGRLFCQPNVYFAQYSGHGVNATHLAGQLLAEAIIGSQSDGFDL 449
Qy 383 --KVPOKIH---ELIKWNPDIANRNW---RDTLG 409
Db 450 FAKVPHITFFGGKLLR--SPILLALGMWYRLKEKLG 483

RESULT 10
US-09-252-991A-30872
; Sequence 30872, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30872
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30872

Query Match 5.4%; Score 125; DB 4; Length 652;
Best Local Similarity 21.8%; Pred. No. 0.00065;
Matches 92; Conservative 54; Mismatches 162; Indels 114; Gaps 16;

Qy 5 SSSLLIVGAGTGTSTALHLAGRGYTNVTVDLP--YPVPSAISAGNDYNNKVISSQYNNK 63
Db 283 NAEVIVVGAGIVGSACAHELARRG--LDVLVDSRRGGATAVGMGHLV-----AMDDNP 334
Qy 64 DEIEVNEILAEAFNGKNDPLFKPYHYDGTLLMSAC-----SQEGDLRLGVVR 113
Db 335 AELALSD--YSIQAWRTWAADLPEDCAYRNCGTWLAADAAELAEAEKRRQALLAAGV--- 390
Qy 114 PGEDPNLVELTRPEQFRKLAPGVQLQDFFQWKGYFARSAGAGWAHARNALVAAAREAQRM 173
Db 391 -----ACEMLDAARLRLDEP--VLR---PGLAGALKVPDGGILYAPN---AARWLLERA 436
Qy 174 GVKFVTGTQGRVVTLIPENNVDKGA--VTGDKGIWFAERTFLCAGASAGQFLDFKNQLR 231
Db 437 G-----PRLRLHAHYSEVDGSRRLADGRWLSAEALVLANGTHAGEL----- 479
Qy 232 PTAWTLVHIALKPEERALKYKNIPVFNIERGFFPEPEERGEIKICDEHP----- 281
Db 480 -----CAELFIRP-----KKGHLITDRTYPTGLRHQLVEL 509
Qy 282 GYTNMQSADGTM-----SIPPEKTOIPKEATRVALLK--ETMPQL 323
Db 510 GYVSAHASSGTSVAFNAQPRGTQGVFLGSSRQFDLTDPQGVFVLAERMLRRALDYLPL 569
Qy 324 ADPRFSPARICWCAADTANRFLIDRHPOYHSLVLCGASGRGFKYLPISGNLIVDAMEGK 383
Db 570 AGLNATRAMTGFRAATPDGLLGEHPAQPGMLAVGHEGLGVTTAPGSRALLAAQLFGE 629
Qy 384 VP 385
Db 630 TP 631

RESULT 11
US-09-489-039A-12890
; Sequence 12890, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:

APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 12890
LENGTH: 476
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12890

Query Match 5.3%; Score 123.5; DB 4; Length 476;
Best Local Similarity 22.5%; Pred. No. 0.00056;

Matches 92; Conservative 48; Mismatches 148; Indels 121; Gaps 21;

QY 6 SLLIIVGAGTGTSTALHARRGYTNVTLDPY-----PVPISAISAGNDVNVK 53
DB 73 ADVVIIGGFGSGINTALEAEQITNVVLEARHLGCGTGRNGGQVMAGI--GHDIEAV 130
QY 54 ISSQGYNNKDEIE-----VNEILAEAFNGKNDPLFKPYHDTGLLMSACSQEGLD 106
DB 131 -----KKHVGKEGLETFLKIANLGAGIIRIRIKYNIDADFVPGY---GYL--AYNORQLK 181
QY 107 RLG-----VRVRPGEDPNLVELTRPEQFKLAPEGVLQDPPGKGYFARSGAGWAHAR 160
DB 182 TLQWKEKFAATPDEE---IELYQKEVQVGVSEV-----YCGALKHMGGGQIHSL 231
QY 161 NALVAAREAQRMGVKFTGT-----QGRVVTLLIFENNNDVKAVTGDGKIWAERT 212
DB 232 NMLLSQAQAHSGLVKIFSSPVVEVNYGKQVRVTAM-----GSVKAALLMACD-- 282
QY 213 FLCAGASAGQFLDKNQLRPTAM--TLV-----HIALKPEERALKYKNIPVIFNIERGFPE 266
DB 283 -----SFLNMEPEIYNKTLVTYSQVSTEPSLDELIERISPL----RGAFS 326
QY 267 PDEERGEIKCDHPGYTNMVQASDGTMSIPEKTIQIPKEATRVALLKETMPOLADR 326
DB 327 IRPVINYVTR-----NRLFGSATRF---VEYT--PNDFAAWNRTLLAEVFPYLRDV 376
QY 327 PFSFARICWADTANREFLIDRHPQVHSLVLGGGASGRGFKYLPISGNL 375
DB 377 KIDFA---WGGPNA-----CSAN-----LFPQIGTL 399

RESULT 12
US-09-684-405-2
Sequence 2, Application US/09684405
Patent No. 6770477
GENERAL INFORMATION:
APPLICANT: Dennis J. Slamon
APPLICANT: Juliana J. Oh
TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED GENES
TITLE OF INVENTION: ASSOCIATED WITH HER-2/NEU OVEREXPRESSION
FILE REFERENCE: 30448.79USU1
CURRENT APPLICATION NUMBER: US/09/684,405
CURRENT FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/157,923
PRIOR FILING DATE: 1999-10-06
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 486
TYPE: PRT
ORGANISM: Homo Sapiens
US-09-684-405-2

Query Match 5.3%; Score 123.5; DB 4; Length 486;
Best Local Similarity 20.9%; Pred. No. 0.00058;
Matches 94; Conservative 53; Mismatches 179; Indels 123; Gaps 20;

QY 6 SLLIIVGAGTGTSTALH-----ARRGYTNVTLDPYPVPSAISAGNDVNVK----- 53
DB 64 SDVVIIGGCVGLSVAYWLKLESRRGAIRVLVVERDHTYSOASTGLSVGGICQFSLPE 123
QY 54 -----ISSQGYNNKDEIEVNEILAEAFNGKNDPLFKPYHDTGLLMSACSQEGLD-R 107
DB 124 NIQLSLFSASFLRN-----INEYLAV-----VDAPPLDLRFNPSGYLLILASEKDAAME 172
QY 108 LGVVRPGEDPNLVELTRPEQFKLAP-----RGVLQDPPGKGYFARSGAGWAHARNAL 163
DB 173 SNVKVQROGAK-VSLMSPDQLRNKFPWINTGVALAS-----YGMEDGWFDPWCLL 224
QY 164 VAAAREAQRMGVKFTGT-----TPQGRVVTL-----TFENNNDVK 198
DB 225 QGLRRKVKVSLGVLFCQGEVTRFVSSSQRMILTDDKAVLKRHEHVHKMDRSLEYQPVEC 284
QY 199 A-VTDGKIWAERTPLCAGASAGQFLDKNQLRPTAMTLVHIALKPEERALKYKNIPVIF 257
DB 285 AIVINAAGAWSAQIAAL-AGVGEPP-----PQTLQGTKLPEVPRKRYVY----- 327
QY 258 NIERGFFPEPDEERGEIKICDEHPGY-TNMVQASDGTM-----MSIPFEKTIQIPKEATR 311
DB 328 -----VWHCPQPGLETPLVADTSGAYFRREGLSNYLGGRSPTQECP 371
QY 312 VRALLK-----ETMPOLADRPFSFARI-----CWCA-----DTANREFLIDRHPQVHSL 355
DB 372 DPANLEVDHDFQDKVWPHALRVAPAFETLKVQSAWAGYYDNTFDQGVGVGPHPLVVM 431
QY 356 VLGGASGRGFKYLPISGNLIVD-AMEGK 383
DB 432 YFATPSGHGLQQAAPGIGRAVAEMVLKGR 460

RESULT 13
US-08-391-339-18
Sequence 18, Application US/08391339
Patent No. 5463175
GENERAL INFORMATION:
APPLICANT: Kishore, Ganesh M.
APPLICANT: Barry, Gerard F.
TITLE OF INVENTION: Glycosate Tolerant Plants
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dennis R. Hoerner, Jr.,
ADDRESSEE: Monsanto Co. BB4F,
STREET: 700 Chesterfield Village Parkway
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/391,339
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/156,968
FILING DATE:
APPLICATION NUMBER: US/07/717,370
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Hoerner, Dennis R., Jr.
REGISTRATION NUMBER: 30,914
REFERENCE/DOCKET NUMBER: 38-21(10533)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)537-6099
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:

```
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-391-339-18

Query Match      5.0%; Score 116.5; DB 1; Length 431;
Best Local Similarity 22.7%; Pred. No. 0.0026;
Matches 94; Conservative 36; Mismatches 119; Indels 165; Gaps 22;

QY 10 IVGAGTGTSTALHLARRGTYNTVLDPPYVPSAISAGNDVNKVISSGQYNNKDEIEVN 69
DB 10 IAGAGIVGCTALMLQRRGF-KVTLLDPNPPGEGASFGN----- 47
QY 70 EILAEAFNGKNDPLPKPYHTDTGLMSACSQEGDLRLG-VVRPGEDP----- 118
DB 48 ----AGCFNGSSVVPMSMP-----GNLTSV-PKWLLDPMGPLSIRFGYFPTIMPWLIRFL 97
QY 119 -----NLVELTRPEQFRKLAPEGVL-----OGDFPGWKGY--FARSGA 154
DB 98 LAGRPNKVKEQAKALRNLIKSTVP-LIKSLAEADASHLIRHEGLHTVYRGEADFARDRG 156
QY 155 GWA-----HARNALVAAAREAQ-----MGVKFVTG-----TPQGRVVTLLIF-- 191
DB 157 GWELRLNGVRTQILSA--DALRDFPNLSHAFTKGLIEENGHTINPQG-LVTLLPRRF 213
QY 192 -----ENNVDVGAVTGDGKIWAERTFLCAGASQGLDFKNQLRPTA 234
DB 214 IANGGBFVSARVIGFETEGRAKGITTTNG-VLAVDAAVVAAGAH----- 258
QY 235 WTLVHIALKEERALKNIPIVNIERGFPEDEERG-EIKICDEHPGYTNMVQSADGT 293
DB 259 -----KSLANSLGDDIPL-----DTERGYHIVIANPEAARIPTTDSAGK 298
QY 294 MMSIPPEK-----TQIP--KEAE--TRVRALLKETMPOLADRPFS 329
DB 299 FIATPMEMGLRVAGTVEFAGLTAAPNWKRAHVLYTRARKLLPALAPASSEERYS 352

RESULT 14
US-08-484-274A-18
; Sequence 18, Application US/08484274A
; Patent No. 576760
; GENERAL INFORMATION:
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Barry, Gerard F.
; TITLE OF INVENTION: Glyphosate Tolerant Plants
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,274A
; FILING DATE: 07 June 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT130 38-21(13560)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713)789-2679
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-484-274A-18

Query Match      5.0%; Score 116.5; DB 1; Length 431;
Best Local Similarity 22.7%; Pred. No. 0.0026;
Matches 94; Conservative 36; Mismatches 119; Indels 165; Gaps 22;

QY 10 IVGAGTGTSTALHLARRGTYNTVLDPPYVPSAISAGNDVNKVISSGQYNNKDEIEVN 69
DB 10 IAGAGIVGCTALMLQRRGF-KVTLLDPNPPGEGASFGN----- 47
QY 70 EILAEAFNGKNDPLPKPYHTDTGLMSACSQEGDLRLG-VVRPGEDP----- 118
DB 48 ----AGCFNGSSVVPMSMP-----GNLTSV-PKWLLDPMGPLSIRFGYFPTIMPWLIRFL 97
QY 119 -----NLVELTRPEQFRKLAPEGVL-----OGDFPGWKGY--FARSGA 154
DB 98 LAGRPNKVKEQAKALRNLIKSTVP-LIKSLAEADASHLIRHEGLHTVYRGEADFARDRG 156
QY 155 GWA-----HARNALVAAAREAQ-----MGVKFVTG-----TPQGRVVTLLIF-- 191
DB 157 GWELRLNGVRTQILSA--DALRDFPNLSHAFTKGLIEENGHTINPQG-LVTLLPRRF 213
QY 192 -----ENNVDVGAVTGDGKIWAERTFLCAGASQGLDFKNQLRPTA 234
DB 214 IANGGBFVSARVIGFETEGRAKGITTTNG-VLAVDAAVVAAGAH----- 258
QY 235 WTLVHIALKEERALKNIPIVNIERGFPEDEERG-EIKICDEHPGYTNMVQSADGT 293
DB 259 -----KSLANSLGDDIPL-----DTERGYHIVIANPEAARIPTTDSAGK 298
QY 294 MMSIPPEK-----TQIP--KEAE--TRVRALLKETMPOLADRPFS 329
DB 299 FIATPMEMGLRVAGTVEFAGLTAAPNWKRAHVLYTRARKLLPALAPASSEERYS 352

RESULT 15
US-09-902-540-9986
; Sequence 9986, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9986
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-9986

Query Match      4.9%; Score 115; DB 4; Length 505;
Best Local Similarity 21.9%; Pred. No. 0.0049;
Matches 92; Conservative 54; Mismatches 148; Indels 126; Gaps 24;

QY 2 VTKSSLLIVGAGTGTSTALHLARRGTYNTVLDPPYVPSAISAGNDVNKVISSGQY- 60
DB 115 VPAKAKVVIIGGIMGLALAYNLARAGETDVVVLERGLCAGASGRNGGVGRVHWGTPL 174
QY 61 --NNKDEIEVNEILAE--AFNGW-----KNDFLPKPYHYDHTGLMSACSQEGDLR 107
DB 175 VELAKRSIELMKGFARELGINVWLROGGYIFLAKTAPVAQLERNVSL-----HNR 225
QY 108 LGVRVRPGEDPNLVELTRPEQFRKLAP-----EGVLQDFPGWKGYFARS 152
```


62 NKDEIEVNEILAEAEFNGWKNDPLEFKPYHDTGLLMSACSOEGLDRLGVRVRPGEDPNLV 121


```

FT XX /note= "Encoded by TTC"
PN
XX WO200261119-A1.
XX
XX 08-AUG-2002.
XX
XX 30-JAN-2002; 2002WO-JP000721.
XX
XX 31-JAN-2001; 2001JP-00022953.
PR
XX 16-FEB-2001; 2001JP-00039796.
PR
XX 08-AUG-2001; 2001JP-00240002.
PR
XX (ASAH ) ASahi KASEI KOGYO KK.
PA
XX Kouzuma T, Yoshioka I, Arai M, Sumitani J, Imamura S;
XX
XX WPI; 2002-599854/64.
XX N-PSDB; ABK90475.
XX
XX Compositions for assaying glycoprotein without interference from globulin
PT and ascorbic acid in blood, useful in clinical examination including
PT diagnosis and management of diabetes.
XX
XX Disclosure; Page 63-65; 82pp; Japanese.
XX
XX The invention relates to compositions for assaying glycoprotein with use
CC of protease and enzymes acting on glycoamino acids in the presence of at
CC least 1 of deoxycholic acid or amide, cholic acid amide octyl glucoside,
CC quaternary ammonium salt, quaternary ammonium salt cationic surfactant,
CC concanavalin A and betaine and/or ascorbic oxidase and a buffer without a
CC 4-(2-hydroxyethyl)-1-piperazinyl group. The invention also relates to
CC compositions for assaying albumin containing the protease and enzymes for
CC determining glycoalbumin and a protein-denaturing agent and/or a compound
CC having S-S bond and bromocresol purple for evaluation of glycation
CC proportion of albumin into glycoalbumin with respect to the other
CC albumins measured separately. The compositions are useful in clinical
CC examination including diagnosis and management of diabetes. The
CC compositions can be used in methods which avoid the effects of globulin
CC acting on glycoamino acids. This sequence represents a polypeptide used
CC in the scope of the invention
XX
SQ Sequence 440 AA;

Query Match 49.7%; Score 1156.5; DB 5; Length 440;
Best Local Similarity 51.3%; Pred. No. 6.3e-108;
Matches 223; Conservative 75; Mismatches 130; Indels 7; Gaps 5;

QY 2 VTKSSLLIVGACTWTSTALHARRGYNTVTVLDPPVPSPSAISAGNDVNVKVIS--SGQY 59
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 LTKSQQLIVGGTGCSTALHARRGYNTVTVLDVNRIPSPISAGHDVNVKLAGRLSTAD 63
QY 60 SNNKDIENVNIIAEAFNGWKNDDPLFKPYHDTGLLMSACSEGLDRGLGRVRPDED-P 118
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 SKGDDSDSWKALSYAAAGWLDHPVFPQCHNTGWSVAGSTPKSIKQL-VEDEIGDDID 122
QY 119 NLVELTRPEOFRLKLAPEGVLQGFPGWKGVFARSGAGWAHARNALVAAAREAGRMGVKFV 178
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 QYTPLNTAEDFRKTMPEGILTGTFPGWKGFKYFTGSGWVHARVAMKAAFEESERLGVKI 182
QY 179 TGTPOGRVTVTLIFENNDRVGAVTGDGKIWAERTFLCAGASAGQFLDFKNQLRPTAWTLV 238
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 TGSPEGKVESLIFEDGDVRCAGTADGKEHRADRTILSAGASAEFFLDFENQIQPTAWTLG 242
QY 239 HIALKPEERALYKNIPVFNIEGPFPEDEERGEIKICDEHPGYTNMQSADGTM-MSI 297
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
243 HIQITPEETKLYKNLPLFNINQGFMEPEDDLHQLKMCDEHPGYNVKEKPGSKYPSQSI 302
QY 298 PFPEKTOIPKEAEFRVRLKLTWPKLADRPFSFARICWCADTANREFFLDRHPQYHSLVL 357
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 PFAXHQVPTAEARMKQFLKDIMPQLADRLVHARICWCADTQDRMFLITYHPRHSLVI 362
QY 358 GCGASGRGFKYLPISIGNLIVDAMEGVKVPQKIHELKWNPDIAANRNWRDTLGRFGGNRV 417
```

```

DB 363 ASGDCCTGYKHITSIGKFIQSDCMEGTLEERFAKYWRWRPEKFTFWGKDFLDRFGADDKI 422
QY 418 MDF--HDVKETNVQ 430
DB 423 MDLPKSDVEGWTNIK 437

RESULT 6
ABG30784
ID ABG30784 standard; protein; 440 AA.
XX
AC ABG30784;
XX
XX 05-NOV-2002 (first entry)
XX
XX Fusarium oxysporum polypeptide #3.
XX
XX Glycoamino acid; glycoprotein; deoxycholic acid; deoxycholic amide;
KW cholic acid amide octyl glucoside; quaternary ammonium salt;
KW quaternary ammonium salt cationic surfactant; concanavalin A; betaine;
KW ascorbic oxidase; 4-(2-hydroxyethyl)-1-piperazinyl group; albumin;
KW protease; glycoalbumin; bromocresol purple; glycation; diabetes;
KW ascorbic acid; mutant; mutein.
XX
XX Fusarium oxysporum.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 246 /note= "Encoded by ATG"
FT Misc-difference 372 /note= "Wild-type Lys substituted by Met"
FT Misc-difference 396 /note= "Encoded by TTC"
FT WO200261119-A1.
XX
XX 08-AUG-2002.
XX
XX 30-JAN-2002; 2002WO-JP000721.
XX
XX 31-JAN-2001; 2001JP-00022953.
PR 16-FEB-2001; 2001JP-00039796.
PR 08-AUG-2001; 2001JP-00240002.
XX
XX (ASAH ) ASahi KASEI KOGYO KK.
XX
XX Kouzuma T, Yoshioka I, Arai M, Sumitani J, Imamura S;
XX
XX WPI; 2002-599854/64.
XX N-PSDB; ABK90477.
XX
XX Compositions for assaying glycoprotein without interference from globulin
PT and ascorbic acid in blood, useful in clinical examination including
PT diagnosis and management of diabetes.
XX
XX Disclosure; Page 69-72; 82pp; Japanese.
XX
XX The invention relates to compositions for assaying glycoprotein with use
CC of protease and enzymes acting on glycoamino acids in the presence of at
CC least 1 of deoxycholic acid or amide, cholic acid amide octyl glucoside,
CC quaternary ammonium salt, quaternary ammonium salt cationic surfactant,
CC concanavalin A and betaine and/or ascorbic oxidase and a buffer without a
CC 4-(2-hydroxyethyl)-1-piperazinyl group. The invention also relates to
CC compositions for assaying albumin containing the protease and enzymes for
CC determining glycoalbumin and a protein-denaturing agent and/or a compound
CC having S-S bond and bromocresol purple for evaluation of glycation
CC proportion of albumin into glycoalbumin with respect to the other
CC albumins measured separately. The compositions are useful in clinical
CC examination including diagnosis and management of diabetes. The
CC compositions can be used in methods which avoid the effects of globulin
CC and ascorbic acid components and stabilise enzymes such as protease
```

```
CC acting on glycoamino acids. This sequence represents a polypeptide used
CC in the scope of the invention
XX
SQ Sequence 440 AA;

Query Match          49.4%; Score 1150.5; DB 5; Length 440;
Best Local Similarity 51.0%; Pred. No. 2.6e-107;
Matches 222; Conservative 75; Mismatches 131; Indels 7; Gaps 5;

QY 2 VTKSSLLIVGAGTGTSTALHLARRGYTNVTLDPYVPSAISAGNDVNKVIS--SGQY 59
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 LTKSQILIVGGGTGCGTSTALHLARRGYTNVTLDPYVPSAISAGNDVNKVIS--SGQY 63
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 60 SNNKDEIVNEILAEAFNCGKNDPLFKPYVYHDTGLLMSACSOEGLDRLGVVRPGE-P 118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 SKGDDSDSIWKALSYAAQGLWHPVFPQCHNTGTSVAGSTPKSIKQL-VEDEIGDDID 122
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 119 NLVELTRPEQFRKLAPEGLVQDFFPGWKGYFARSAGAHARNALVAAAREARMGVKFV 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 123 QYTPLNTAEDFRKTMPEGILTGNFPGWKGYKPTGSGVHARKAMKAAFESERLGVKFI 182
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 179 TGTPOQRVVTILFENNNDVKGAVTGDGKIWAERTFLCAGASAGQFLDFKNQLRPTAWTLV 238
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 183 TGSPEGKVESLIFEDGDVGRGAKTADGKEHRADRTILSAGASAEFFLDFENQIQPTAWTLG 242
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 239 HIALKPERALYKNIPIVFNIERGFPEPDEERGEIKICDEHPGYTNMVSADGTM-MSI 297
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 HIQITPEETKLYKNLPLFNINQGFMEPEDDLHLQKMCDEHPGYCNMVKPKSGKYQSI 302
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 298 PFEKTQIPKEATRVRLALKETMPQLADRPFPSPARICWCADTANREFLIDRHPYHSLVL 357
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 PFAKHQVPTAERRMKQFLKDIMPQLADRLVHARICWCADTQDRMFLIYTHRPHSLVI 362
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 358 GCGASGRGFKYLPISGNLIVDAMEGKVPQKIHKLKNPDIANRNWRDTLGRFGGPNRV 417
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 363 ASGDCGTGYVHITSIGKFIKSDCMEGTLEERFAKYWRWRPEKFTFWGKDPDLDRFGADDKI 422
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 418 MDF--HDVKEWTNVQ 430
   |||:|||||:
Db 423 MDLPKSDVEGWNTIK 437
   |||:|||||:

RESULT 7
ABG30785
ID ABG30785 standard; protein; 440 AA.
XX
AC ABG30785;
XX
DT 05-NOV-2002 (first entry)
DE
DE Fusarium oxysporum polypeptide #4.
XX
KW Glycoamino acid; glycoprotein; deoxycholic acid; deoxycholic amide;
KW cholic acid amide octyl glucoside; quaternary ammonium salt;
KW quaternary ammonium salt cationic surfactant; concanavalin A; betaine;
KW ascorbic oxidase; 4-(2-hydroxyethyl)-1-piperazinyl group; albumin;
KW protease; glycoalbumin; bromocresol purple; glycation; diabetes;
KW ascorbic acid; mutant; muten.
XX
OS Fusarium oxysporum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 246 /note= "Encoded by ATG"
FT FT
FT Misc-difference 372 /note= "Wild-type Lys substituted by Val"
FT FT
FT Misc-difference 396 /note= "Encoded by TTC"
FT FT
XX WO200261119-A1.
PN
PN 08-AUG-2002.
PD
```

```
XX 30-JAN-2002; 2002WO-JP000721.
XX
XX 31-JAN-2001; 2001JP-00022953.
PR 16-FEB-2001; 2001JP-00039796.
PR 08-AUG-2001; 2001JP-00240002.
XX
PA (ASAH ) ASAH KASEI KOGYO KK.
XX
PI Kouzuma T, Yoshioka I, Arai M, Sumitani J, Imamura S;
XX WPI; 2002-599854/64.
XX N-PSDB; ABK90478.
XX
PT Compositions for assaying glycoprotein without interference from globulin
PT and ascorbic acid in blood, useful in clinical examination including
PT diagnosis and management of diabetes.
XX
PS Disclosure; Page 72-75; 82pp; Japanese.
XX
CC The invention relates to compositions for assaying glycoprotein with use
CC of protease and enzymes acting on glycoamino acids in the presence of at
CC least 1 of deoxycholic acid or amide, cholic acid amide octyl glucoside,
CC quaternary ammonium salt, quaternary ammonium salt cationic surfactant,
CC concanavalin A and betaine and/or ascorbic oxidase and a buffer without a
CC 4-(2-hydroxyethyl)-1-piperazinyl group. The invention also relates to
CC compositions for assaying albumin containing the protease and enzymes for
CC determining glycoalbumin and a protein-denaturing agent and/or a compound
CC having S-S bond and bromocresol purple for evaluation of glycation
CC proportion of albumin into glycoalbumin with respect to the other
CC albumins measured separately. The compositions are useful in clinical
CC examination including diagnosis and management of diabetes. The
CC compositions can be used in methods which avoid the effects of globulin
CC and ascorbic acid components and stabilise enzymes such as protease
CC acting on glycoamino acids. This sequence represents a polypeptide used
CC in the scope of the invention
XX
SQ Sequence 440 AA;

Query Match          49.4%; Score 1149.5; DB 5; Length 440;
Best Local Similarity 51.0%; Pred. No. 3.2e-107;
Matches 222; Conservative 75; Mismatches 131; Indels 7; Gaps 5;

QY 2 VTKSSLLIVGAGTGTSTALHLARRGYTNVTLDPYVPSAISAGNDVNKVIS--SGQY 59
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 LTKSQILIVGGGTGCGTSTALHLARRGYTNVTLDPYVPSAISAGNDVNKVIS--SGQY 63
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 60 SNNKDEIVNEILAEAFNCGKNDPLFKPYVYHDTGLLMSACSOEGLDRLGVVRPGE-P 118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 SKGDDSDSIWKALSYAAQGLWHPVFPQCHNTGTSVAGSTPKSIKQL-VEDEIGDDID 122
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 119 NLVELTRPEQFRKLAPEGLVQDFFPGWKGYFARSAGAHARNALVAAAREARMGVKFV 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 123 QYTPLNTAEDFRKTMPEGILTGNFPGWKGYKPTGSGVHARKAMKAAFESERLGVKFI 182
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 179 TGTPOQRVVTILFENNNDVKGAVTGDGKIWAERTFLCAGASAGQFLDFKNQLRPTAWTLV 238
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 183 TGSPEGKVESLIFEDGDVGRGAKTADGKEHRADRTILSAGASAEFFLDFENQIQPTAWTLG 242
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 239 HIALKPERALYKNIPIVFNIERGFPEPDEERGEIKICDEHPGYTNMVSADGTM-MSI 297
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 HIQITPEETKLYKNLPLFNINQGFMEPEDDLHLQKMCDEHPGYCNMVKPKSGKYQSI 302
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 298 PFEKTQIPKEATRVRLALKETMPQLADRPFPSPARICWCADTANREFLIDRHPYHSLVL 357
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 PFAKHQVPTAERRMKQFLKDIMPQLADRLVHARICWCADTQDRMFLIYTHRPHSLVI 362
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 358 GCGASGRGFKYLPISGNLIVDAMEGKVPQKIHKLKNPDIANRNWRDTLGRFGGPNRV 417
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 363 ASGDCGTGYVHITSIGKFIKSDCMEGTLEERFAKYWRWRPEKFTFWGKDPDLDRFGADDKI 422
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 418 MDF--HDVKEWTNVQ 430
   |||:|||||:
Db 423 MDLPKSDVEGWNTIK 437
   |||:|||||:
```

Db 423 MDLPKSDVEGWNIK 437

RESULT 8

ABG30783

ID ABG30783 standard; protein; 440 AA.

XX

AC ABG30783;

XX

DT 05-NOV-2002 (first entry)

XX

DE Fusarium oxysporum polypeptide #2.

XX

KW Glycoamino acid; glycoprotein; deoxycholic acid; deoxycholic amide;

KW cholic acid amide octyl glucoside; quaternary ammonium salt;

KW quaternary ammonium salt cationic surfactant; concanavalin A; betaine;

KW ascorbic oxidase; 4-(2-hydroxyethyl)-1-piperazinyl group; albumin;

KW protease; glycoalbumin; bromocresol purple; glycation; diabetes;

KW ascorbic acid; mutant; mutein.

XX

OS Fusarium oxysporum.

OS Synthetic.

XX

XX

FT Key Location/Qualifiers

FT Misc-difference 246 /note= "Encoded by ATG"

FT Misc-difference 372 /note= "Wild-type Lys substituted by Trp. Encoded by TTG"

FT Misc-difference 396 /note= "Encoded by TTC"

FT

XX WO20026119-A1.

XX

XX 08-AUG-2002.

XX

PF 30-JAN-2002; 2002WO-JP000721.

XX

XX 31-JAN-2001; 2001JP-00022953.

PR 16-FEB-2001; 2001JP-00039796.

PR 08-AUG-2001; 2001JP-00240002.

XX

PA (ASAH) ASahi KASEI KOGYO KK.

XX

PI Kouzuma T, Yoshioka I, Arai M, Sumitani J, Imamura S;

XX

XX WPI: 2002-599854/64.

DR N-PSDB; ABK30476.

XX

XX Compositions for assaying glycoprotein without interference from globulin and ascorbic acid in blood, useful in clinical examination including diagnosis and management of diabetes.

PT

XX Disclosure; Page 66-69; 82pp; Japanese.

XX

CC The invention relates to compositions for assaying glycoprotein with use of protease and enzymes acting on glycoamino acids in the presence of at least 1 of deoxycholic acid or amide, cholic acid amide octyl glucoside, quaternary ammonium salt, quaternary ammonium salt cationic surfactant, concanavalin A and betaine and/or ascorbic oxidase and a buffer without a 4-(2-hydroxyethyl)-1-piperazinyl group. The invention also relates to compositions for assaying albumin containing the protease and enzymes for determining glycoalbumin and a protein-denaturing agent and/or a compound having S-S bond and bromocresol purple for evaluation of glycation proportion of albumin into glycoalbumin with respect to the other albumins measured separately. The compositions are useful in clinical examination including diagnosis and management of diabetes. The compositions can be used in methods which avoid the effects of globulin and ascorbic acid components and stabilise enzymes such as protease acting on glycoamino acids. This sequence represents a polypeptide used in the scope of the invention

XX

SQ Sequence 440 AA;

Query Match 49.3%; Score 1148.5; DB 5; Length 440;

Best Local Similarity 51.0%; Pred. No. 4.1e-107;

Matches 222; Conservative 75; Mismatches 131; Indels 7; Gaps 5;

QY 2 VTKSSLLIVGAGTGTSTALHLARRGTYNTVLDPPVPSAISAGNDVNVKVIS--SGQY 59

Db 4 LTKSQIILIVGGTGWGCTSTALHLARRGTYNTVLDVNRIPSPISAGHDVNVKLAGRLSTAD 63

QY 60 SNNKDEIEVNEITLAEAEAFNGWKNDFLPKPYHYDHTGLMSACSQEGLDRLDGVVRPGED-P 118

Db 64 SKGDDSDSIWKALSYAAAGQWLHDFVFPQCHTGVVAGSTPKSIKQL-VEDEIGDDID 122

QY 119 NULVELTRBEQFKLAPBGLVQDPPGKGYFARSAGAHARNALVAAAREAQRMGVKFV 178

Db 123 QYTPLNTAEDFRKTMPEGILTCNFPCKWKFYKPTGSGWVHARKAMKAAFEESERLGVKFI 182

QY 179 TCTPGRVVTVLIFENNVDKGVATGDKIWAERTFLCAGASAGQFLDKNQLRPTAWTLV 238

Db 183 TGSPGKVESLIFEDGDVRGAKTAGKEHRAORTILSAGASAEFFLDFFENQIQPTAWTLG 242

QY 239 HIALKPEERALYKNIPIFNIERGFFFEDEBERGEIKICDEHPGYTNMVQSADGTW-MGI 297

Db 243 HIQITPEETKLYKNLPPLFNINQGFMEDEDLHQLKMCDEHPGYCNWVEKPSKYQSI 302

QY 298 PFEKTIQPKAEATRVALLKETMPOLADRPFSFARIWCADTANREFLIDRHPQYHSLVL 357

Db 303 PPAKHQVPTAEARRMKQFLKDIMPOLADRLVHARICWCADTQDRMFLITYHPRHPSLVI 362

QY 358 GCGASGRGPKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANNRNRDITLGRFGGPNRV 417

Db 363 ASGDCGTGYWHITSIGKFI SDCMEGLEERPAKYWRWRPEKTEFWGKDPDLDRFGADDKI 422

QY 418 MDF--HDVKEWTNVQ 430

Db 423 MDLPKSDVEGWNIK 437

RESULT 9

ADM78789

ID ADM78789 standard; protein; 444 AA.

XX

AC ADM78789;

XX

DT 01-JUL-2004 (first entry)

XX

DE Fusarium proliferatum fructosylamine oxidase Q2 protein sequence.

XX

XX fructosylamine oxidase; fructosylllysine; fructosyllvaline;

KW Amadori compound; medical examination; diagnosis; serum glucose level;

KW food inspection; enzyme.

XX

OS Fusarium proliferatum.

XX

PN WO2004029251-A1.

XX

PD 08-APR-2004.

XX

XX 16-SEP-2003; 2003WO-JP011766.

XX

PR 24-SEP-2002; 2002JP-00277214.

PR 24-OCT-2002; 2002JP-00309734.

XX

PA (ARKR-) ARKRAY INC.

XX

XX Yoshida N, Tani Y, Yonehara S;

XX WPI: 2004-316127/29.

DR N-PSDB; ADM78788.

XX

XX Fusarium proliferatum-originated fructosylamine oxidase useful in measurement of Amadori compound and particularly applicable in medical examination including diagnosis and management of diabetes.

PT

XX

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:28:55 ; Search time 2.63539 Seconds

(without alignments)
2331.700 Million cell updates/sec

Title: US-10-622-893A-1

Perfect score: 63

Sequence: 1 MGSGDDDLAL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	48	76.2	64	033246	Q33246 mycobacteri
2	48	76.2	64	Q7T212	Q7T212 mycobacteri
3	47	74.6	230	Q6H4X1	Q6H4X1 oryza sativ
4	46	73.0	105	Q6ZE21	Q6ZE21 oryza sativ
5	46	73.0	392	Q6FXT3	Q6FXT3 candida gla
6	46	73.0	399	Q9NE98	Q9NE98 leishmania
7	46	73.0	951	Q6FQ85	Q6FQ85 candida gla
8	46	73.0	1039	Q9N974	Q9N974 leishmania
9	45	71.4	586	Q6PLR1	Q6PLR1 triticum ae
10	44	69.8	182	Q8L432	Q8L432 oryza sativ
11	44	69.8	240	Q23686	Q23686 arabidopsis
12	44	69.8	359	Q6Z134	Q6Z134 oryza sativ
13	44	69.8	621	Q24X11	Q24X11 arabidopsis
14	44	69.8	652	Q7XRMO	Q7XRMO oryza sativ
15	44	69.8	718	Q1YTR1	Q1YTR1 tupaiid her
16	44	69.8	758	Q8SAX7	Q8SAX7 oryza sativ
17	44	69.8	997	Q6K7F3	Q6K7F3 oryza sativ
18	43	68.3	216	Q6EPB1	Q6EPB1 oryza sativ
19	43	68.3	300	Q84P83	Q84P83 oryza sativ
20	43	68.3	316	Q69X10	Q69X10 oryza sativ
21	43	68.3	333	Q9ZS77	Q9ZS77 hordeum vul
22	43	68.3	344	Q8RYL3	Q8RYL3 oryza sativ
23	43	68.3	361	Q6Z726	Q6Z726 oryza sativ
24	43	68.3	512	Q873G5	Q873G5 neurospora
25	43	68.3	652	Q6YU05	Q6YU05 oryza sativ
26	43	68.3	708	Q8S555	Q8S555 picea abies
27	43	68.3	894	Q9U297	Q9U297 caenorhabdi
28	43	68.3	925	Q86JQ8	Q86JQ8 dictyosteli
29	42	66.7	64	Q30517	Q30517 mycobacteri
30	42	66.7	70	Q8GVR6	Q8GVR6 oryza sativ
31	42	66.7	90	Q9BY99	Q9BY99 homo sapien

32	42	66.7	94	2	Q6K896	Q6K896 oryza sativ
33	42	66.7	118	2	Q6UW77	Q6UW77 oryza sativ
34	42	66.7	135	2	Q69WT7	Q69WT7 oryza sativ
35	42	66.7	142	2	Q7PWL9	Q7PWL9 anopheles g
36	42	66.7	203	2	Q7XND1	Q7XND1 oryza sativ
37	42	66.7	214	2	Q8RYU4	Q8RYU4 oryza sativ
38	42	66.7	215	2	Q6TX11	Q6TX11 rattus norv
39	42	66.7	313	2	Q18704	Q18704 endocrypau
40	42	66.7	316	2	Q6FTB7	Q6FTB7 candida gla
41	42	66.7	341	2	Q7FLJ7	Q7FLJ7 oryza sativ
42	42	66.7	353	2	Q96XY4	Q96XY4 sulfolobus
43	42	66.7	417	2	Q8GRJ1	Q8GRJ1 oryza sativ
44	42	66.7	455	2	Q6BGW1	Q6BGW1 debaryomyce
45	42	66.7	484	2	Q84Z25	Q84Z25 oryza sativ

ALIGNMENTS

RESULT 1

ID	Q33246	PRELIMINARY;	PRT;	64 AA.
AC	Q33246; Q7D711;			
DT	01-JAN-1998 (TrEMBLrel. 05, Created)			
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)			
DT	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)			
DE	Hypothetical protein.			
GN	OrderedLocusNames=MT2171, Rv2111c;			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterinae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230; DOI=10.1038/311159;			
RA	Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C.M.,			
RA	Harris D.E., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III,			
RA	Tekaia F., Badcock K., Basham D., Brown D., Chillingworth T.,			
RA	Connor R., Davies R.M., Delvin K., Feltwell T., Gentles S., Hamlin N.,			
RA	Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S.,			
RA	Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,			
RA	Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RX	MEDLINE=22206494; PubMed=12218036;			
RX	DOI=10.1128/JB.184.19.5479-5490.2002;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,			
RA	Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,			
RA	Salzberg S.L., Delcher A., Utterback T.R., Weidman J.F., Khouiri H.M.,			
RA	Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,			
RA	Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	J. Bacteriol. 184:5479-5490 (2002).			
DR	EMBL; BX842578; CAB10702.1; -.			
DR	EMBL; AE000516; AAK46454.1; -.			
DR	PIR; B70512; B70512.			
DR	TIGR; MT2171; -.			
DR	TubercuList; Rv2111c; -.			
DR	InterPro; IPR008515; DUF797.			
DR	Pfam; PF05639; DUF797; 1.			
SQ	SEQUENCE 64 AA; 6944 MW; 7D958B770D8766E3 CRC64;			

Query Match 76.2%; Score 48; DB 2; Length 64;

Best Local Similarity 80.0%; Pred. No. 1.1;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY      2 GGGGDDDDLA 11
Db      10 GGGGDDDDIA 19

RESULT 2
Q7T1212
ID Q7T1212 PRELIMINARY; PRT; 64 AA.
AC Q7T1212;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein Mb2135C.
GN OrderedLocusNames=Mb2135C;
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248341; CAD96988.1; -.
DR InterPro; IPR008515; DUF797.
DR Pfam; PF05639; DUF797; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 64 AA; 6944 MW; 7D958B770D8766E3 CRC64;

Query Match 76.2%; Score 48; DB 2; Length 64;
Best Local Similarity 80.0%; Pred. No. 1.1;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 GGGGDDDDLA 11
Db      10 GGGGDDDDIA 19

RESULT 3
Q6H4X1
ID Q6H4X1 PRELIMINARY; PRT; 230 AA.
AC Q6H4X1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein OSJNBa0085K21.17 (Hypothetical protein
DE OSJNBa0023117.33).
GN Names=OSJNBa0085K21.17; Synonyms=OSJNBa0023117.33;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248341; CAD96988.1; -.
DR InterPro; IPR008515; DUF797.
DR Pfam; PF05639; DUF797; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 64 AA; 6944 MW; 7D958B770D8766E3 CRC64;

Query Match 76.2%; Score 48; DB 2; Length 64;
Best Local Similarity 80.0%; Pred. No. 1.1;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 GGGGDDDDLA 11
Db      10 GGGGDDDDIA 19

RESULT 4
Q6ZE21
ID Q6ZE21 PRELIMINARY; PRT; 105 AA.
AC Q6ZE21;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein P0495H05.26.
GN Name=P0495H05.26;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248341; CAD96988.1; -.
DR InterPro; IPR008515; DUF797.
DR Pfam; PF05639; DUF797; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 105 AA; 11527 MW; DE2D43D9C75D50CF CRC64;

Query Match 73.0%; Score 46; DB 2; Length 105;
Best Local Similarity 72.7%; Pred. No. 4.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 MGGSGDDDDLA 11
Db      69 LGGGGDDDDSA 79

RESULT 5
Q6FXT3
ID Q6FXT3 PRELIMINARY; PRT; 392 AA.
AC Q6FXT3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Similar to sp|P34223|Saccharomyces cerevisiae YBL058w SHP1.
GN ORFNames=CAGL0A04345g;
OS Candida glabrata CBS138.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=284593;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS138;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrans P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neugeglise C., Talia E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekalia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
```

```

RL Nature 430:35-44(2004).
DR EMBL; CR380947; CAG57849.1; -.
DR InterPro; IPR006556; DUF_PAF1.
DR InterPro; IPR009060; UBA_like.
DR PFAM; PF00789; UBX; 1.
DR SMART; SM00553; SEP; 1.
DR SMART; SM00166; UBX; 1.
DR PROSITE; PS00033; UBX; 1.
SQ SEQUENCE 392 AA; 43786 MW; 80FF82FCB5E398B CRC64;

Query Match 73.0%; Score 46; DB 2; Length 392;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGGDDDD 9
Db 106 GSGGDDDD 113

RESULT 6
Q9NE98 PRELIMINARY; PRT; 399 AA.
AC Q9NE98
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein L4803.06;
GN Names=L4803.06;
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Masuy D., Purnelle B., Goffeau A., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL161416; CAB77684.1; -.
KW Hypothetical protein.
SQ SEQUENCE 399 AA; 43458 MW; BB86021C2BBB3E18 CRC64;

Query Match 73.0%; Score 46; DB 2; Length 399;
Best Local Similarity 88.9%; Pred. No. 20;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGGDDDD 10
Db 255 GSGGDDDEL 263

RESULT 7
Q6FQ85 PRELIMINARY; PRT; 951 AA.
AC Q6FQ85
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Candida glabrata strain CBS138 chromosome I complete sequence.
GN ORFNames=CAGL0108305g;
OS Candida glabrata CBS138.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=284593;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=CBS138;
RG Genotelevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Karest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenne D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Solotkin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpetti C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CR380955; CAG60546.1; -.
DR GO; GO:0008415; Fcylitransferase activity; IEA.
DR InterPro; IPR000542; Carn_acyl_trans.
DR PFAM; PF00755; Carn_acyltransf; 1.
SQ SEQUENCE 951 AA; 106462 MW; FB5334D57409D4E1 CRC64;

Query Match 73.0%; Score 46; DB 2; Length 951;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGGSGDDDDIAL 12
Db 899 IGDEGDDDDIAL 910

RESULT 8
Q9N974 PRELIMINARY; PRT; 1039 AA.
ID Q9N974
AC Q9N974
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein P1295.14;
GN Name=P1295.14;
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Masuy D., Purnelle B., Goffeau A., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL359773; CAB95222.1; -.
DR HSSP; Q92830; If68.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011124; Znf_CW.
DR PFAM; PF07496; zf-CW; 1.
KW Hypothetical protein.
SQ SEQUENCE 1039 AA; 109546 MW; 3A91F5A70147E6B3 CRC64;

Query Match 73.0%; Score 46; DB 2; Length 1039;
Best Local Similarity 72.7%; Pred. No. 63;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSGGDDDDIAL 12

```

```

Db      841 GGGGDDDDVPL 851
||| |||||: |
RESULT 9
Q6PLR1 PRELIMINARY; PRT; 586 AA.
AC Q6PLR1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Polyphenol oxidase (Fragment).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RA Jukanti A.K., Bruckner P.L., Fischer A.M.;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY596266; AAT0523.1; -.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR008922; Di-copper centre.
DR InterPro; IPR000896; Hemocyanin.
DR InterPro; IPR006311; Tat.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00264; Tyrosinase; 1.
DR PRINTS; PR00092; TYROSINASE.
DR TIGRFAMs; TIGR01409; TAT signal seq. 1.
DR PROSITE; PS00210; HEMOCYANIN 2; UNKNOWN_1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
FT NON_TER 1
SQ SEQUENCE 586 AA; 63631 MW; D4A1E80D26EE33FE CRC64;

Query Match 71.4%; Score 45; DB 2; Length 586;
Best Local Similarity 80.0%; Pred. No. 47;
Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 3 GGGDDDDDLAL 12
|:|||||
Db 36 GTGDDDDVL 45

RESULT 10
Q8L432 PRELIMINARY; PRT; 182 AA.
AC Q8L432;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein OSJNB0086108.9 (Hypothetical protein
OSJNB0086108.4).
GN ORFNames=OSJNB0086108.4, OSJNB0086108.9;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Kim M.M.,
RA Overton II L.L., Bera J.J., Tsitrin T., Krol M.I., Jarrahi B.B.,
RA Jin S.S., Koo H., Zismann V., Heiao J., Blunt S., Vanaken S.S.,
RA Utterback T.T., Feldblyum T.V., Yang Q.Q., Haas B.J., Suh B.B.,
RA Peterson J.J., Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Buell R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RN [3]

```

```

RP SEQUENCE FROM N.A.
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RT chromosome 10.";
RL Science 300:1566-1569 (2003).
RN [4]
RP SEQUENCE FROM N.A.
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC105746; RAM92808.1; -.
DR EMBL; AC105746; RAM92811.1; -.
DR EMBL; AE017066; AAP52498.1; -.
DR EMBL; AE017066; AAP52503.1; -.
DR Gramene; O8L432; -.
KW Hypothetical protein.
SQ SEQUENCE 182 AA; 19521 MW; BDD9842B26EB3F9D CRC64;

Query Match 69.8%; Score 44; DB 2; Length 182;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCGSGDDDDLAL 12
|||:||||: |
Db 1 MCGAGGDDDRVEL 12

RESULT 11
O23686 PRELIMINARY; PRT; 240 AA.
ID O23686
AC O23686;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein At2g30480.
GN Name=At2g30480;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Town C.D., Haas B.J., Wu D., Maiti R., Hannick L.I., Chan A.P.,
RA Tallon L.J., Rooney T., Utterback T.R., Vanaken S.E., Feldblyum T.V.,
RA White O., Fraser C.M.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Town C.D., Kaul S.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U93215; AABG3096.1; -.
DR PIR; H84708; H84708.
KW Hypothetical protein.
SQ SEQUENCE 240 AA; 27400 MW; A6AB9BAF4E1CF6CC CRC64;

Query Match 69.8%; Score 44; DB 2; Length 240;
Best Local Similarity 77.8%; Pred. No. 25;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCGSGDDDD 9
|||:|||||
Db 163 VGGTGDDDD 171

RESULT 12
Q6Z134 PRELIMINARY; PRT; 359 AA.
ID Q6Z134
AC Q6Z134;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)

```

```

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Hypothetical protein P0455F03.29.
GN Name=P0455F03.29;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
CLoned:P0455F03.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005454; BAC84337.1; -.
DR InterPro; IPR004949; DUF266.
DR Pfam; PF03267; DUF266; 1.
KW Hypothetical protein.
SQ SEQUENCE 359 AA; 39100 MW; C9FDD923301906CB CRC64;

Query Match 69.8%; Score 44; DB 2; Length 359;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MGGSGDDDDLLAL 12
Db 44 VGGGGEDGDLAL 55

RESULT 13
Q84X11 ID Q84X11 PRELIMINARY; PRT; 621 AA.
AC Q84X11;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein At2g30480/T6B20.25;
GN Name=At2g30480/T6B20.25;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Xiao Y.L., Malik M., Whitelaw C.A., Town C.D.;
RX MEDLINE=22369114; PubMed=12481096; DOI=10.1104/pp.010207;
RA Xiao Y.L., Malik M., Whitelaw C.A., Town C.D.;
RT "Cloning and sequencing of cDNAs for hypothetical genes from
chromosome 2 of Arabidopsis.";
RL Plant Physiol. 130:2118-2128(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Xiao Y., Smith S.R., Ishmael N., Kumar N., Redman J., Riedmuller S.,
RA Uterback T., Whitelaw C.A., Fraser C.M., Town C.D.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY219097; AAO37184.1; -.
KW Hypothetical protein.
SQ SEQUENCE 621 AA; 71079 MW; A38A7DE170AFC9D3 CRC64;

Query Match 69.8%; Score 44; DB 2; Length 621;
Best Local Similarity 77.8%; Pred. No. 75;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGGSGDDDD 9
Db 490 VGGGGDDDD 498

RESULT 14
Q7XRMO ID Q7XRMO PRELIMINARY; PRT; 652 AA.
AC Q7XRMO;

```

```

DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSJNBa0058G03.4 protein.
GN Name=OSJNBa0058G03.4;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA PubMed=12447439; DOI=10.1038/nature01183;
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Wang Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.;
RT "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
DR EMBL; AL731606; CAB02429.1; -.
DR Gramene; Q7XRMO; -.
SQ SEQUENCE 652 AA; 72720 MW; DF79F807D9962BD4 CRC64;

Query Match 69.8%; Score 44; DB 2; Length 652;
Best Local Similarity 80.0%; Pred. No. 80;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GGS GGDDDDLLA 11
Db 21 GGS GGDDDDRA 30

RESULT 15
Q91TRI ID Q91TRI PRELIMINARY; PRT; 718 AA.
AC Q91TRI;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE T32.
OS Tupaiid herpesvirus 1 (strain 1) (TuHV-1) (Herpesvirus tupaia (strain
1)).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae.
OX NCBI_TaxID=10397;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=2;
RX MEDLINE=21211637; PubMed=11312357;
RX DOI=10.1128/JVI.75.10.4854-4870.2001;
RA Bahr U., Darai G.;
RT "Analysis and characterization of the complete genome of tupaia (tree
shrew) herpesvirus.";
RL J. Virol. 75:4854-4870(2001).
DR EMBL; AF281817; AAK57076.1; -.
DR Pfam; PF06070; Herpes_UL32; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN 1.
SQ SEQUENCE 718 AA; 74589 MW; 92ED97DAC14EA58F CRC64;

Query Match 69.8%; Score 44; DB 2; Length 718;
Best Local Similarity 88.9%; Pred. No. 89;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GGS GGDDDDLL 10
Db 374 GGLGGDDDDLL 382

```

Search completed: May 27, 2005, 13:47:24
Job time : 6.63539 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:47:36 ; Search time 126.721 Seconds
(without alignments)
1189.648 Million cell updates/sec

Title: US-10-622-893A-3

Perfect score: 2329

Sequence: 1 AVTKSSLLIVGAGTGTST.....MDFHVKWNTVQYRDISKL 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1462099 seqs, 344972447 residues

Total number of hits satisfying chosen parameters: 1462099

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2329	100.0	437	17 US-10-622-893A-3	Sequence 3, Appli
2	2329	100.0	472	17 US-10-622-893A-5	Sequence 5, Appli
3	1151.5	49.4	440	17 US-10-470-678-9	Sequence 9, Appli
4	1150.5	49.4	440	17 US-10-470-678-10	Sequence 10, Appli
5	1149.5	49.4	440	17 US-10-470-678-8	Sequence 8, Appli
6	625	26.8	437	14 US-10-232-655-3	Sequence 3, Appli
7	600	25.8	437	14 US-10-232-655-1	Sequence 1, Appli
8	210.5	9.0	139	16 US-10-767-701-53616	Sequence 53616, A
9	184	7.9	384	14 US-10-156-761-14478	Sequence 14478, A
10	166	7.1	387	17 US-10-829-427-1	Sequence 1, Appli
11	165	7.1	385	15 US-10-282-122A-68225	Sequence 68225, A
12	160	6.9	433	14 US-10-287-274-412	Sequence 412, App
13	160	6.9	433	15 US-10-282-122A-42701	Sequence 42701, A

14	159.5	6.8	381	14	US-10-156-761-9328	Sequence 9328, Ap
15	140	6.0	374	15	US-10-282-122A-55515	Sequence 55515, A
16	136.5	5.9	372	15	US-10-282-122A-43213	Sequence 43213, A
17	135.5	5.8	378	10	US-09-975-719-367	Sequence 367, App
18	135	5.8	420	15	US-10-425-114-54552	Sequence 54552, A
19	127	5.5	372	15	US-10-282-122A-75876	Sequence 75876, A
20	126.5	5.4	433	9	US-09-815-242-13745	Sequence 13745, A
21	126	5.4	370	15	US-10-282-122A-72847	Sequence 72847, A
22	124	5.3	429	15	US-10-425-114-71622	Sequence 71622, A
23	123	5.3	414	16	US-10-437-963-177242	Sequence 177242, A
24	122	5.2	406	17	US-10-485-710-80	Sequence 80, Appli
25	119.5	5.1	391	15	US-10-282-122A-69762	Sequence 69762, A
26	119.5	5.1	818	14	US-10-156-761-14479	Sequence 14479, A
27	113.5	4.9	395	9	US-09-873-880-30	Sequence 30, Appli
28	113.5	4.9	395	15	US-10-424-599-268694	Sequence 268694, A
29	113.5	4.9	852	15	US-10-282-122A-50546	Sequence 50546, A
30	112.5	4.8	317	15	US-10-282-122A-61710	Sequence 61710, A
31	112.5	4.8	405	14	US-10-156-761-9148	Sequence 9148, Ap
32	112.5	4.8	542	16	US-10-437-963-117091	Sequence 117091, A
33	112.5	4.8	626	15	US-10-447-013-27	Sequence 27, Appli
34	112.5	4.8	629	9	US-09-741-669-473	Sequence 473, App
35	112.5	4.8	629	9	US-09-815-242-10405	Sequence 10405, A
36	112.5	4.8	629	15	US-10-282-122A-56758	Sequence 56758, A
37	112	4.8	563	15	US-10-415-302-8	Sequence 8, Appli
38	112	4.8	563	15	US-10-415-302-18	Sequence 18, Appli
39	110.5	4.7	412	9	US-10-416-537-2	Sequence 2, Appli
40	110.5	4.7	632	15	US-09-873-880-32	Sequence 32, Appli
41	110.5	4.7	632	15	US-10-282-122A-68951	Sequence 68951, A
42	110.5	4.7	797	15	US-10-369-493-9773	Sequence 9773, Ap
43	109.5	4.7	621	15	US-10-282-122A-55707	Sequence 55707, A
44	109	4.7	481	16	US-10-437-963-168785	Sequence 168785, A
45	108.5	4.7	837	15	US-10-282-122A-49792	Sequence 49792, A

ALIGNMENTS

RESULT 1
US-10-622-893A-3
; Sequence 3, Application US/10622893A
; Publication No. US20050014935A1
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Sheng
; APPLICANT: Datta, Abhijit
; APPLICANT: Wang, Yuping
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; DETERMINATION OF GLYCATED PROTEINS
; FILE REFERENCE: 46692001300
; CURRENT APPLICATION NUMBER: US/10/622,893A
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 40%-100% identity to the amadoriase
US-10-622-893A-3

Query Match	100.0%	Score	2329	DB	17	Length	437
Best Local Similarity	100.0%	Pred. No.	3.7e-226				
Matches	437	Conservative	0	Mismatches	0	Indels	0
		Gaps	0				
Qy	1	AVTKSSLLIVGAGTGTSTALHARGYNTVLDYPVPSAISAGNDVKNVLSGQYS	60				
Db	1	AVTKSSLLIVGAGTGTSTALHARGYNTVLDYPVPSAISAGNDVKNVLSGQYS	60				
Qy	61	NNKDEIVNEILAEAEFNGWKNDPLPKFPYVHDTGLLMSACSQEGLDLRLGVVRVPEDPNL	120				
Db	61	NNKDEIVNEILAEAEFNGWKNDPLPKFPYVHDTGLLMSACSQEGLDLRLGVVRVPEDPNL	120				
Qy	121	VELTRPEQFKLAPEGVLOQDFFGWKGYFARSGAGWAHARNALVAAAREAQRMGVKFTVTG	180				

Db 121 VELTRPEQFRKLAPEGVLQGDFFGWKGYFARSAGAGWAHARNALVAAAREAQRMGVKFTVG 180
QY 181 TPQGRVVTLLIFENNVDKVGAVTGDGKIWRAERTFLCAGASAGQFLDFKNQLRPTAWTLVHI 240
Db 181 TPQGRVVTLLIFENNVDKVGAVTGDGKIWRAERTFLCAGASAGQFLDFKNQLRPTAWTLVHI 240
QY 241 ALKPEERALYKNIPIVFNIERGFFFPDEERGEIKICDEHPGYTNMVQSADGTMSIPPE 300
Db 241 ALKPEERALYKNIPIVFNIERGFFFPDEERGEIKICDEHPGYTNMVQSADGTMSIPPE 300
QY 301 KTOIPKEAETRVALLKETMPOLADRPSPFARICWCADTANREFLIDRHPQYHSLVLCG 360
Db 301 KTOIPKEAETRVALLKETMPOLADRPSPFARICWCADTANREFLIDRHPQYHSLVLCG 360
QY 361 ASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANRNWRDTLGRFGPNRVMDF 420
Db 361 ASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANRNWRDTLGRFGPNRVMDF 420
QY 421 HDVKWNTNVQYRDISKL 437
Db 421 HDVKWNTNVQYRDISKL 437

RESULT 2

US-10-622-893A-5
; Sequence 5, Application US/10622893A
; Publication No. US20050014935A1
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Sheng
; APPLICANT: Datta, Abhijit
; APPLICANT: Wang, Yuping
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; DETERMINATION OF GLYCATED PROTEINS
; FILE REFERENCE: 466992001300
; CURRENT APPLICATION NUMBER: US/10/622,893A
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric protein
US-10-622-893A-5

Query Match 100.0%; Score 2329; DB 17; Length 472;
Best Local Similarity 100.0%; Pred. No. 4.2e-226;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVTKSSLLIVGAGTGTSTALHLARRGYTNVTLDPPYVPSAISAGNDVNKVISGGQYS 60
Db 13 AVTKSSLLIVGAGTGTSTALHLARRGYTNVTLDPPYVPSAISAGNDVNKVISGGQYS 72
QY 61 NNKDEIVNEILAEAFNGKNDPLFKPYHYDHTGLLMSACSQSGLDRLGVRVRPGEDPDL 120
Db 73 NNKDEIVNEILAEAFNGKNDPLFKPYHYDHTGLLMSACSQSGLDRLGVRVRPGEDPDL 132
QY 121 VELTRPEQFRKLAPEGVLQGDFFGWKGYFARSAGAGWAHARNALVAAAREAQRMGVKFTVG 180
Db 133 VELTRPEQFRKLAPEGVLQGDFFGWKGYFARSAGAGWAHARNALVAAAREAQRMGVKFTVG 192
QY 181 TPQGRVVTLLIFENNVDKVGAVTGDGKIWRAERTFLCAGASAGQFLDFKNQLRPTAWTLVHI 240
Db 193 TPQGRVVTLLIFENNVDKVGAVTGDGKIWRAERTFLCAGASAGQFLDFKNQLRPTAWTLVHI 252
QY 241 ALKPEERALYKNIPIVFNIERGFFFPDEERGEIKICDEHPGYTNMVQSADGTMSIPPE 300
Db 253 ALKPEERALYKNIPIVFNIERGFFFPDEERGEIKICDEHPGYTNMVQSADGTMSIPPE 312
QY 301 KTOIPKEAETRVALLKETMPOLADRPSPFARICWCADTANREFLIDRHPQYHSLVLCG 360

Db 313 KTOIPKEAETRVALLKETMPOLADRPSPFARICWCADTANREFLIDRHPQYHSLVLCG 372
QY 361 ASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANRNWRDTLGRFGPNRVMDF 420
Db 373 ASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANRNWRDTLGRFGPNRVMDF 432
QY 421 HDVKWNTNVQYRDISKL 437
Db 433 HDVKWNTNVQYRDISKL 449

RESULT 3

US-10-470-678-9
; Sequence 9, Application US/10470678
; Publication No. US20050101771A1
; GENERAL INFORMATION:
; APPLICANT: KOUZIMA, Takuji et al.
; TITLE OF INVENTION: COMPOSITION FOR ASSAYING GLYCATED PROTEINS
; FILE REFERENCE: 1516-0121P
; CURRENT APPLICATION NUMBER: US/10/470,678
; CURRENT FILING DATE: 2003-07-30
; PRIOR APPLICATION NUMBER: JP 2001/ 22953
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: JP 2001/ 39796
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: JP 2001/ 24002
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic protein sequence of SEQ ID NO: 3
; NAME/KEY: MUTAGEN
; LOCATION: (372)..(372)
US-10-470-678-9

Query Match 49.4%; Score 1151.5; DB 17; Length 440;
Best Local Similarity 51.0%; Pred. No. 4.2e-107;
Matches 222; Conservative 75; Mismatches 131; Indels 7; Gaps 5;
QY 2 VTKSSLLIVGAGTGTSTALHLARRGYTNVTLDPPYVPSAISAGNDVNKVIS--SGQY 59
Db 4 LTKSQQLIVGGTGWGCTALHLARRGYTNVTLDVNRIPSPISAGHDVKNLAGRLSTAD 63
QY 60 SNNKDEIVNEILAEAFNGKNDPLFKPYHYDHTGLLMSACSQSGLDRLGVRVRPGED-P 118
Db 64 SKGDDSDSIWKALSVAQAQGLHDPVFPQFCHNTGWSVAGSTPKSIKQL-VEDEIGDDID 122
QY 119 NLVELTRPEQFRKLAPEGVLQGDFFGWKGYFARSAGAGWAHARNALVAAAREAQRMGVK 178
Db 123 QYTPLNTAEDFRKTMPEGILTGNFPGWKGYFPTGSGWVHARKAMKAAFEESRLGVKFI 182
QY 179 TGTPOGRVVTLLIFENNVDKVGAVTGDGKIWRAERTFLCAGASAGQFLDFKNQLRPTAWTLV 238
Db 183 TGSPEGVESLIFEDGDVGRGAKTADGKEHADRTILSAGASAEFFLDFENQIQTAWTLG 242
QY 239 HIALKPEERALYKNIPIVFNIERGFFFPDEERGEIKICDEHPGYTNMVQSADGTMSI 297
Db 243 HIQMTPEETKLYKNLPFLFNINQGFMEPEDELHQLKMCDEHPGYNWVEKPGSKYPQSI 302
QY 298 PEKTOIPKEAETRVALLKETMPOLADRPSPFARICWCADTANREFLIDRHPQYHSLV 357
Db 303 PFAKHQVTEAERMRKQFLKIMPOLADRLVHARICWCADTODRWFLLTYHHPHPSLVI 362
QY 358 GCASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANRNWRDTLGRFGPNRV 417
Db 363 ASGDCGTGYMHTSICKFISDCMEGLTEERFAKFWWRPEKFTFPGKOPDLRFGADDKI 422
QY 418 MDF--HDVKWNTNVQ 430


```

; SEQ ID NO 1
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Coniochaeta sp.
US-10-232-655-1

Query Match      25.8%; Score 600; DB 14; Length 437;
Best Local Similarity 32.7%; Pred. No. 2.4e-51;
Matches 141; Conservative 86; Mismatches 162; Indels 42; Gaps 13;

QY      8 LLIVGAGTCTGTALHLARGY--TNTVLDPPVPSAISAGNDVNKVVISGQYSNNKDE 65
DB      11 IVVGGGTIGSSSTALHLVRSGYAPANTVLDTFEIPSAQSAGHDUNKIM--GIRLRNKVD 68

QY      66 IEVNEIIAEAFNGKNDPLFKPYTHDTGLIMSACSOEGLDRLGVRVRPGEDPNL-VELT 124
DB      69 LQ-----MSLEARQWKEDELFPFFHNTGRMDCHEHTPEGIEDLKKQYQALHDAGAGLEKT 124

QY      125 R---PEQFRKLAPGVLOGD-FPGWKGYFARSGAGWAHARNALVAAAREAGRMGVKFTVG 180
DB      125 HAWLDNDEILSKMPLLRDQIQGWKAIWSOG-GWLAAAAKAIINAIGQFLKRGYKFGFG 183

QY      181 TPQGRVVTLLFENNVDK--GAVTGDGKIWAERTFLCAGASAGQFLDPKNOLRPTAWTLV 238
DB      184 -GAGSFQKPLFDEGTTTCIGVETADGTYKIVADKVVLAAGAWSPTLVDLDEQCCSAWYA 242

QY      239 HIAKPERALYKNIPIVFNIERGFFPEOBERGEIKICDSEHPGYTNMVQ-----SADGT 293
DB      243 HIQLTPAEAAEYKGVVYNGEGFFPEPN-EFGVIKVCDEFFGFSRPFKEHPYCAPSPK 301

QY      294 MMSIPFEKTOIPKE-----AETRVRLALKETMPQLADRPFSFARICWCADTANREFLIDR 348
DB      302 HISVPRSHAKHPTDTPDASEVSIKKAIAITFLPRFQDKELFNRLCWCCTDTADAALLMCE 361

QY      349 HPQYHSLVLCGASGRGFKYLPSCNLIVDAMEGKVQKIHILKNWP----- 396
DB      362 HPKWNKFIATGDSGHSFKILPNVGKHVVELIGRLPEEMAYQWRWRPGGDALKSRRAAP 421

QY      397 --DIAANNRWR 405
DB      422 PKLADMPGHW 432

RESULT 8
US-10-767-701-53616
; Sequence 53616, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 53616
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 13394327.pcp
US-10-767-701-53616

Query Match      9.0%; Score 210.5; DB 16; Length 139;
Best Local Similarity 33.6%; Pred. No. 9.7e-13;
Matches 44; Conservative 21; Mismatches 39; Indels 27; Gaps 3;

QY      303 QIPKEAETRVRLALKETMPQLADRPFSFARICWCADTANREFLIDRHYPQYHSLVLCGAS 362
DB      7 EIPKEGDVCRFLRECIPEGNRPWTHTRICWYTDTPGDMVIDYHPKYGKGLFVATGGS 66

```

Qy	363	GRGFKYLPISGNI	LYVDAMEGKVPQKI	HELIIKWNPDIA	NRNRDTLG-----	RFGGPNRV	417
Db	67	GHGYKEVPKIGRI	VDVMK-----	QPRDDLGTETFK	QSWPQK	106	
Qy	418	MDPHDYKEWTN	428				
Db	107	YHFDHI--WTN	115				

RESULT 9

```

US-10-156-761-14478
; Sequence 14478, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE NUMBER: 249-262
; CURRENT APPLICATION NUMBER: US/10156,761
; CURRENT FILING DATE: 2002-05-29
; PRIORITY DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIORITY DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14478
; LENGTH: 384
; TYPE: Prt
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14478

```

Query Match	7.9%;	Score 184;	DB 14;	Length 384;
Best Local Similarity	22.1%;	Pred. No. 2.2e-09;		
Matches	101;	Conservative 60;	Mismatches 139;	Gaps 158;
QY	2	VTKSSSLLIVCAGTWCSTALHLARRGYTNVTVL-----	35	
		: : : : : : : : : : : : : : : :		
Db	1	VSPYDVIVILGGMGSAHAHLISARG-----ARVLGLEKFGPVHNRGSSHGSRITROS YF	57	
		: : : : : : : : : : : : : : : :		
QY	36	-DPVPVPSAI-----SAGDNVKNVISSGOYSNNKDEIEVNEILAEAFNGWQND	83	
		: : : : : : : : : : : : : : : :		
Db	58	EDPAYVPLLRSYELYEVEVSTGREV-ATLSGGVMVGRPDSLTV-----	101	
		: : : : : : : : : : : : : : : :		
QY	84	PLFKPYHYDGLLMSACS-----QEGDLRLGVRVR-PGEDPNLVELTRPEQPKLAPEGVL	138	
		: : : : : : : : : : : : : : : :		
Db	102	-----AGLSRSATQWDLPEHMLDAKEIRRRFPTLNPSENDEALVEK-----	142	
		: : : : : : : : : : : : : : : :		
QY	139	QGDFPGWKGYPARSGAGWAHARNALVAAAREAOBMGVKFTGTGPGQVRVVTLLIFENNVDVKG	198	
		: : : : : : : : : : : : : : : :		
Db	143	-----KAGLVRPENNVAAHLQLATRQGAEE-----LHFEPETRW	176	
		: : : : : : : : : : : : : : : :		
QY	199	AVTGDG-KIWAERTFLCAGASAGQFLDFKNQLRPTATWLVIHALKPEERALLYKNIPIVF	257	
		: : : : : : : : : : : : : : : :		
Db	177	EPYRDGVRVHTAENTY-----TAGQLV-----ICFGAWA-----PQ-----LLTDLGVFPF	216	
		: : : : : : : : : : : : : : : :		
QY	258	NIERG-----PFPEDEERGEIKCDHPGTVMNVQSDGTM-----SIPPEKT	302	
		: : : : : : : : : : : : : : : :		
Db	217	TIERQWYWFQPHRGVGPFR-PENHPYI--IWEDAEGVQVYGFPSIDGPDLGAKVAFPK	273	
		: : : : : : : : : : : : : : : :		
QY	303	QIPKEAFT-----RVRLAKKETMPLQADRPFSF--ARICWCADTANREFLIHRFPQY	352	
		: : : : : : : : : : : : : : : :		
Db	274	GVVCTPETIDRTVHDVHQAMADHMSRCIPDLPGTFLKAATCMY/SNTPEHFVFIARPAH	333	
		: : : : : : : : : : : : : : : :		
QY	353	-HSLVLCCGASGRGFKYLPISGNLIVD-AMEGKVPQKI	388	
		: : : : : : : : : : : : : : : :		
Db	334	PDSVTVACGSGHGKFKVPVVGVEIVADLALGTGTAHPI	371	
		: : : : : : : : : : : : : : : :		

RESULT 10

US-10-829-427-1
; Sequence 1, Application US/10829427
; Publication No. US20050026265A1
; GENERAL INFORMATION:
; APPLICANT: FURUKAWA, KEISUKE
; APPLICANT: KAJIYAMA, NAOKI
; TITLE OF INVENTION: MODIFIED SARCOSINE OXIDASES, GENES AND RECOMBINANT DNAs THEREOF,
; TITLE OF INVENTION: AND METHODS FOR PREPARING THE SAME
; FILE REFERENCE: 252020USO
; CURRENT APPLICATION NUMBER: US/10/829,427
; CURRENT FILING DATE: 2004-04-22
; PRIOR APPLICATION NUMBER: JP 2003-121533
; PRIOR FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: JP 2003-396807
; PRIOR FILING DATE: 2003-11-27
; PRIOR APPLICATION NUMBER: JP 2004-116345
; PRIOR FILING DATE: 2004-04-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-829-427-1

```

Query Match          7.1%; Score 166; DB 17; Length 387;
Best Local Similarity 21.4%; Pred. No. 1.5e-07;
Matches      89; Conservative 68; Mismatches 173; Indels 86; Gaps 20;

QY      8 LLTVGAGTGTCTALHLARRGYNTVTLPDPYPVPSAISAGNDVNKVYSSQQYSNNKDEIE 67
       :::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      7 VVVVGAGSGMGAAGYYLAKQGVKTLLV-DAPDPPHTEGSHHGDTRIIRHA-YGEGRKYVP 64
       ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY     68 VNEILABEAFNGKNDPLPKPYHYHDGLL-----MSAQOEGL-----D 106
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB     65 F-ALRAQELWYLENETHNK-IFTKTVLVFGPKGSDFAETMEAAAEHSLTVDLLEGD 122
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY    107 RLCGRVRPGEDPNVLVELTRPEQFRKLAPGSVLOQDFPQWKGVFARSAGWAHARNALVA 166
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   123 EINTR-WPG-----ITPVENYNAI-----FEPNSGVLFSENCIRS 157
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY   167 AREAQRMGVKFVTGTPOGRVVTLLIFENNDDVKGAVTDGDKIWRRAERTFLCAGASAGQLDF 226
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   158 RELAVAQAKILTYT---RVEDFEVSQDV-K-IQTANGS-YTADKLIVSMGAWNSKLSK 212
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY   227 KQOLRPTANTLVHIAKP-----ERALYKN---IPV-IFNIERGEFFPEPDERGE 273
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   213 LN-----LDLPQRYRVVGFSPFDNSNAKSYNDVGYPAPMVVEYPMKGIYTGFPFSGG- 262
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY   274 IKICDBHPGVTNWVSQADGMWSISPEKTKQIPKEAETRVRALLKETMPOLADRPPSFARI 333
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   263 ---CGLKIGHYHYGQIDPDPTINREGAVQ---EDESNLDRDLEKXWPE-ANGELKRGA 315
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY   334 CMCADTANREFIDRHPOYHSIVLVCAGSAGRKFYLPISIGNLIVD-AMEGKVPQKI 308
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   316 CMYTKTPDHFFVIDTHPEHSNVAAAGFSGHGFKFSSVVGEVLSQLATTGKTHEDI 371
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 11
US-10-282-122A-68225
; Sequence 68225, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel

```

RESULT 11

```

US-10-282-122A-69225
; Sequence 69225, Application US/10282122A
; Publication NO. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Halsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel

```

```
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68225
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Pseudomonas putida
US-10-282-122A-68225
```

```
Query Match 7.1%; Score 165; DB 15; Length 385;
Best Local Similarity 23.5%; Pred. No. 1.8e-07;
Matches 101; Conservative 53; Mismatches 161; Indels 114; Gaps 22;

QY 4 KSSLLIVAGTGTSTALHLARRGVTNTVLDPPVPSAISAGNDVNKVIS-----56
Db 6 KSYTAVVGVGVGSATLWELAAQ---QQDVL-----GLEAGAPINLQSSYGGSRIF 55

QY 57 -GOYSNNKDEIEVNEILAEAFNGWK--NDPLFKPYHTDTGLLMSACSGEGLDRLGVRVR 113
Db 56 RQAYWEGSDYLS---LLAEDLGMRELQATSHRPLLYSG-----GLFTG 97

QY 114 PGEDPNLVELTRPEQFRKLAPEGVLQDPPGKGYFARGAGWAHARNALVAAAREA---170
Db 98 P-----IRSGVSVGSAAS-----AKAG-GIAHOR--LTAAEVBARFS 131

QY 171 -----QRMGVKFVTGT-----PQGRVVTLLIFENNNDVKGA---TG 202
Db 132 VFRADENMEAVFQGAFTIAADARLQMLNQAHAHGQMRFGSHVQDITRVESEFLLRLS 191

QY 203 DGKIWAERTFLCAGAS-AGQFL-DFKNQLRPTAWTLVHIALKPERALYKNIPIV-IFNI 259
Db 192 DQGSVLAQKVVLATGAGLAGSLIPDLISGLLRPSRVPYIYCAFKSGAGQLFANPAFLYEL 251

QY 260 ERG-FFFFEEDERGE--IKICDEHPCYTNMVOSA-DGTMMSIPPEKTOIPKEAETRVRA 314
Db 252 EDGRLLYGTPOIDNAEPGIKI-----GFHNQOSALDPTQLEPASDAQIEE-----ISA 301

QY 315 LLKETWPLQADRPFSFARICWCADTANREFLIDRHPQYHSLVLCGASGRGPKYLSIGN 374
Db 302 CVSRVPDPLIARYA-SRKCVTYMTPEAFIIGESKELPSVFFVCSGSGHGFKFAALGS 360

QY 375 LIVDAMEGK 383
Db 361 CLARALAGQ 369
```

RESULT 12

```
US-10-287-274-412
; Sequence 412, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THEREOF
; FILE REFERENCE: ELITRA.008DVL
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 412
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-287-274-412
```

```
Query Match 6.9%; Score 160; DB 14; Length 433;
Best Local Similarity 23.3%; Pred. No. 7.1e-07;
Matches 106; Conservative 49; Mismatches 146; Indels 154; Gaps 24;

QY 8 LLIVAGTGTSTALHLARRGVTNTVLDPPVPSAISAGNDVNKVISGQYNNKDETE 67
Db 17 IIIIGAGTAGTACALRCARAG-LSVLLLERAEIPG-----SKNLSGRL-----59

QY 68 VNEILAEAFNGWKNDPLFKPYHTDTGLLMSACSGEGLDRL---GVRVTPGEDPNLVELT 124
Db 60 YTHALAE-----LLPQFHLTAPLERRITHESUSLLTPDGVT-----96
```

```
QY 125 RPEQFRKLAPEG---VLQDPPGKGYFARGAGWAHARNALVAAAREAQMVKFVTG 180
Db 97 ----FSSLPQGGSWSVLARFDPW-----LVA---EAEKEGVECI---130

QY 181 TPQGRVVTLLIFENNNDVKGAVTGDGKIWAERTFLCAGASAGQFLDFKNQLRPTAWTLVHI 240
Db 131 -PGATVDALYEENGRVCGVICGD-DILRARYVVLAEAG-----NSVLAERHGLV--177

QY 241 ALKPERALYKNIPIVFN-----ISRGFFEPDEERGEI---KICDEHPG---YTNMVQ 288
Db 178 -TRPAGEAMALGIKEVLSLETSABERPHLENNEGALLPFSGRICDDLPGGAPLYTNQOT 236

QY 289 SADG-----TMMSIPPEKTOIPKEAETRVALLKET-----MPQ--LADRPFS 329
Db 237 LSLGIVCPSSLTQSRVPASELLTRFKAHPAVRPLIKNTESLEYGAHLVPEGGLHSMVPQ 296

QY 330 FARICW-----CADTANREFLIDRHPQYHSLVLCGASGRGPKY----LPSTGNLI 376
Db 297 YAGNGWLLVGDALRSCVNT-----GISVRGMDMALTCQAQAAQTL 336

QY 377 VDAMEGVPOKTHELIKKNPDIANENWEDTLGRF 411
Db 337 ISACQHRPEQNLPPL--YHHNVERSLW-DVLQRY 368
```

RESULT 13

```
US-10-282-122A-42701
; Sequence 42701, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: May 27, 2005, 13:34:21 ; Search time 33.5437 Seconds
(without alignments)
1253.491 Million cell updates/sec

Title: US-10-622-893A-3

Perfect score: 2329

Sequence: 1 AVTKSSLLIVGAGTGTST.....MDPHDVKEWTVQYRDISKL 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 79:**

2: PIR1:**

3: PIR2:**

4: PIR3:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	488.5	21.0	412	2 T40295	fructosyl amine -
2	465	20.0	433	2 T37605	probable fructosyl
3	192	8.2	390	2 I39975	sarcosine oxidase
4	181	7.8	389	2 JS0671	sarcosine oxidase
5	177.5	7.6	390	2 JC7256	L-pipecolate oxida
6	168	7.2	389	2 T44248	sarcosine oxidase
7	160	6.9	433	1 B65058	fixC protein homol
8	155	6.7	387	2 JU0461	sarcosine oxidase
9	150	6.4	433	2 E91081	hypothetical prote
10	150	6.4	433	2 F85926	hypothetical prote
11	148	6.4	380	2 AB3160	santhopine deamina
12	146.5	6.3	427	2 G83298	conserved hypothet
13	142.5	6.1	857	2 S16133	dimethylglycine de
14	139.5	6.0	364	2 B83078	probable D-amino a
15	138	5.9	376	2 A83905	hypothetical prote
16	137.5	5.9	372	2 A85668	sarcosine oxidase-
17	137.5	5.9	372	2 E90808	sarcosine oxidase-
18	136.5	5.9	372	2 JC5371	probable sarcosine
19	136.5	5.9	382	2 G75122	sarcosine oxidase,
20	136.5	5.9	416	2 AD3075	sarcosine oxidase
21	136.5	5.9	416	2 E98211	sarcosine oxidase
22	135.5	5.8	372	2 F90159	sarcosine oxidase,
23	134.5	5.8	825	2 F95963	probable dehydroge
24	132	5.7	371	2 F83487	hypothetical prote
25	128.5	5.5	382	2 F71008	probable sarcosine
26	127	5.5	372	2 A10637	opine catabolism h
27	127	5.5	372	2 A70019	probable oxidoredu
28	126.5	5.4	433	2 A10882	thiamin biosynthes
29	126	5.4	652	2 AH2245	

30	125.5	5.4	417	2 G98200	sarcosine oxidase
31	125.5	5.4	417	2 AB3086	sarcosine oxidase
32	125.5	5.4	439	2 E82983	probable oxidoredu
33	125.5	5.4	442	2 D97382	hypothetical prote
34	125.5	5.4	442	2 AC2600	oxidoreductase ord
35	123.5	5.3	437	2 T45533	agaE protein (impo
36	116.5	5.0	452	2 G95356	probable oxidoredu
37	114	4.9	396	2 AF2835	sarcosine oxidase
38	114	4.9	413	2 A97613	sarcosine oxidase
39	113	4.9	420	2 AC3284	D-amino-acid dehyd
40	113	4.9	703	2 T15503	hypothetical prote
41	112.5	4.8	629	1 BVECOA	glucose inhibited
42	112.5	4.8	629	2 C91214	glucose-inhibited
43	112.5	4.8	629	2 D86060	glucose-inhibited
44	112	4.8	405	2 I40887	sarcosine oxidase
45	111.5	4.8	441	2 F98306	agaE protein (impo

ALIGNMENTS

RESULT 1

T40295

fructosyl amine - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C;Accession: T40295

R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Hilbert, H.; Duesterhoeft, A.

submitted to the EMBL Data Library, March 1998

A;Reference number: Z21918

A;Accession: T40295

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-412 <MOO>

A;Cross-references: UNIPROT:O43029; EMBL:AL022071; PIDN:CAA17815.1; GSPDB:GN00067; SPDB:

A;Experimental source: strain 972h-; cosmid c354

C;Genetics:

A;Gene: SPDB:SPBC354.15

A;Map position: 2

Query Match	21.0%;	Score 488.5;	DB 2;	Length 412;
Best Local Similarity	30.1%;	Pred. No. 5e-32;		
Matches	130;	Conservative	78;	Mismatches 183; Indels 41; Gaps 12;
Qy	2	VTKSSLLIVGAGTGTSTALHLARRGYTNVTVLDPYVPSAISAGNDVNKVISSGOYSN	61	
Db	1	MVKNTSVIIVGAGVFGLSAALELTGRGYTIKILDRAPPVIDGSSVDANRIIRS-DYAD	59	
Qy	62	NKDEIEVNEILAEAFNGWKNDPLFKPYHDTGLMSA-----CSQEGDLRLGYRVR	113	
Db	60	-----AVYCSMGIDALEENRTNLFKEQFYGSGLMFVGRDNVEYRDMSENLTOMGV---	111	
Qy	114	PGEDPNLVELTRPEQPKLAPEGVLQDPFGWGYFARSGAGWAHARNALVAAREAQRM	173	
Db	112	-----SAAKPQTTEERLKLFPKWI--GELNDGAGYANFSSGWAHAQSVKSVVNYLAHA	164	
Qy	174	GKVFVTGTPQGRVTVLIFENNVDKGAVTGDKGIWRAERTFLCAGASAGQL-DFKNQLRP	232	
Db	165	GVSFISG-PBGTVEELITEENVVKGVRTTG-AVMAEKLIFATGAWTASLLPNDRFLA	222	
Qy	233	TATLVHIALKPERALYKNIPIVNIERGFPEPDEERGEIKCDHPGYTNMVQSGADG	292	
Db	223	TGQFVAYIKLTPEYIRFLTNPVYLPDFTGTGYIFPPPTPDGLYLFARHGYGFTRMNLKSG	282	
Qy	293	TWMSIPEKTQI-----PKAEATRVALLKETM-POLADRPFSFARICWCADTANREFLID	347	
Db	283	KVESVPPKPLVSLPKEAELOLRNLQRTYEEISQRFYKTRICYTDTADAEVFD	342	
Qy	348	RHPQYSHLVLCGASGRGFKYLPISIGNLIVDAMEGKVPQKIHILIKW---NPDIANRNW	404	
Db	343	YHPDYENLNVCTGSGHGFFKFFPLGKYISGCMFRELEELPKKWRKKNLEPAALDHS	402	
Qy	405	RDTLGRFGGPNR	416	

Db 403 R-----AGPSR 408 ||:|

RESULT 2
T37605
probable fructosyl amino acid oxidase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T37605
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Saunders, D.; Harris, D.
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21729
A:Accession: T37605
A:Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: DNA
A:Residues: 1-433 <CD>
A:Cross-references: UNIPROT:Q9UTW9; EMBL:AL132667; PIDN:CAB59618.1; GSPDB:GN00066; SPDB:
A:Experimental source: strain 972h-; cosmid c139
C:Genetics:
A:Gene: SPDB:SPAC139.04C
A:Map position: 1

Query Match 20.0%; Score 465; DB 2; Length 433;
Best Local Similarity 30.1%; Pred. No. 4.5e-30;
Matches 130; Conservative 80; Mismatches 174; Indels 48; Gaps 13;

QY 5 SSSLLLVGAGTGTSTALHLAR-RGYTNVTVDLPYPVPSAISAGNDVNKVISSGOYSNNK 63
Db 2 SRTIVVGGCVGLSTAVELAKNHSFDNTIAIDAEVPSPSSMAANDINKIVRP-EYADLK 60
QY 64 DEIEVNEILAAEFNWKNDPLFKPYHDTGLLMSACSOEGLDRLGVRVRPGE----- 116
Db 61 -----YMKLAAEMEKWRNDPELSSVYFECGRIST-----TSKDPYRARFDEVAQRNL 109
QY 117 -----DNLVELTRPEQFRKLAP-----EGVLQDGFPGWKGYFARSGAGWAHARNALVAAR 168
Db 110 KLLGDSALINLSSEIRKKYPLSPNSPLRSDMQA-----VNEHAGYANSAASKLEL 165
QY 169 EAQRMGVKFTGTPQGR-----VV-----TLFENNVDKGAVTGDKIWAERTFLCAGASA 220
Db 166 KARELGVEFVFG-KAGFKFKFVNHSSETDIDKNDNHVSQVTEGDTIYHADTILLAVGAYL 224
QY 221 GQFLDFKNQLRPTAWTLVHIALKPERALYKNIPVFNIERGFFPEPDEERGEIKICDEH 280
Db 225 NAYLNTSHRCAGLPAHQLTDEBFKYKNPILFPDPCAFAFPYPVTKLIKLASTG 284
QY 281 PGYTNMVQS---ADGTWMSIPF---EKTQIPKEAETRVLRALLKETMPQLADRPFSFARIC 334
Db 285 YEYVCNVETDYDENSKVSIPIHSGPSKSLPKYALIQMRFLDTFLDPLADRLSLINTKMC 344
QY 335 WCADTANREFLIDRHPQYHSLVLCGASGRGFKYLSIGNLIYDVAMEGKVPOKIHILIKW 394
Db 345 WISDTEADNPLDKVPQDNVFNVDGSDGHAFKFLPNIGRYIAQRIILGDLSEWKDAWRW 404
QY 395 NPDIYAANR-NWR 405
Db 405 REDDKASELKWR 416

RESULT 3
I39975
sarcosine oxidase (EC 1.5.3.1) - Bacillus sp.
C:Species: Bacillus sp.
C>Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 09-Jul-2004
C:Accession: I39975
R:Suzuki, K.; Inamura, S.; Sugiyama, M.
J. Ferment. Bioeng. 77, 231-234, 1994
A:Title: Cloning, sequencing, overexpression in Escherihia coil of a sarcosine oxidase-
A:Reference number: I39975
A:Accession: I39975
A:Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: DNA

A:Residues: 1-390 <RES>
A:Cross-references: UNIPROT:P40859; GB:D16521; NID:G984787; PIDN:BAA03967.1; PID:G984788
C:Genetics:
A:Gene: sox
C:Superfamily: sarcosine oxidase
C:Keywords: oxidoreductase

Query Match 8.2%; Score 192; DB 2; Length 390;
Best Local Similarity 23.2%; Pred. No. 8.1e-08;
Matches 97; Conservative 68; Mismatches 162; Indels 92; Gaps 22;

QY 8 LLIVGAGTGTSTALHLARRGTYTNVTVDLPYPVPSAISAGNDVNKVISSGOYSNNKDEIE 67
Db 7 VIVVGAGSMGAAGYQLAKQGVKTLIV-DAFPPHTNGSHHGDTIRHA-YEGREYVP 64
QY 68 VNEILAAEFNWKNDPLFKPYH---DTGLLMSACSOEGLDRLGVRVRPQGEDPNLVEL 123
Db 65 -----LALRSQELWYE--LEKETHHKIFTKTGLV-----FGPK---GESAFVAET 105
QY 124 TRPEQFRKLAPGVLQGD-----FPGW---KGVFA--RSGAGWAHARNALVAAAREAQRM 173
Db 106 MEAAKEHSLTVD-LLEGDEINKRWPGITVPENYNALFEFNSGVLFSENCIRAYRELAER 164
QY 174 GVKFTVTGTPQGRVVTLLIFENNVDKGAVTGDKIWAERTFLCAGASAGQFLDFKNQLRPT 233
Db 165 GAKVLTHT---RVEDFDISPDSVK-IETANGS-YTADKLIVSMGAWNSKLLSKLN----- 214
QY 234 ANTLVHIALKPERALYKNIPVFNIERGFFPEPDEERGEIKICDEHPGYTNMVQSADGT 293
Db 215 -----LDIPLQ-----YRQV-----GPFESDESKYSNDI--DFFPGF--MVEVPNGI 253
QY 294 MMSIPE-----EKTQIPKEAETRVLRALLKETMPQLADRPFSF 330
Db 254 YGFPFSGCGGLKLGVHTFGQKIDPDTINREFGVPEDESNLRAFLEETWPG-ANGELKR 312
QY 331 ARICWCADTANREFLIDRHPQYHSLVLCGASGRGFKYLPSPISGNLIYD-AMEGKVPOKI 388
Db 313 GAVCMYTKLDEHFIDLHPEHSNVVIAAGFSGHGFKFSSGVGEVLSQALTGKTSHDI 371

RESULT 4
JS0671
sarcosine oxidase (EC 1.5.3.1) precursor - Streptomyces sp.
C:Species: Streptomyces sp.
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 20-Jun-2000
C:Accession: JS0671; PS0345
R:Suzuki, K.; Ogishima, M.; Sugiyama, M.; Inouye, Y.; Nakamura, S.; Inamura, S.
BioSci. Biotechnol. Biochem. 56, 432-436, 1992
A:Title: Molecular cloning and expression of a Streptomyces sarcosine oxidase gene in Str
A:Reference number: JS0671; MUID:92330009; PMID:1368326
A:Accession: JS0671
A:Molecule type: DNA
A:Residues: 1-389 <SUZ>
A:Cross-references: GB:D10623; NID:G217023; PIDN:BAA01473.1; PID:G217024
A:Experimental source: strain KB210-8S
A:Accession: PS0345
A:Molecule type: protein
A:Residues: 2-31;122-143;230-259;268-283 <SUZ1>
A:Note: the authors translated the initiation codon GTG for residue 1 as Val
C:Comment: This enzyme catalyzes the oxidation of sarcosine to form formaldehyde, glycine
C:Superfamily: sarcosine oxidase
C:Keywords: oxidoreductase
F:2-389/Product: sarcosine oxidase #status experimental <SAR>
F:11-16/Region: nucleotide binding #status predicted

Query Match 7.8%; Score 181; DB 2; Length 389;
Best Local Similarity 22.9%; Pred. No. 6.4e-07;
Matches 100; Conservative 59; Mismatches 150; Indels 128; Gaps 24;

QY 8 LLIVGAGTGTSTALHLARRGTYTNVTVDLPYPVPSAISAGNDVNKVISSGOYSNNKDEIE 67
Db 7 VIVIGLGMGSAAAHLSARG-ARVLGLEKF-----GPNHNRSSHG 47

RESULT 7

B65058
fixC protein homolog b2766 - Escherichia coli (strain K-12)
N:Contains: probable quinone reductase (EC 1.6.5.-)
C:Species: Escherichia coli
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002
C:Accession: B65058
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A6720; MUID:97426617; PMID:9278503
A:Accession: B65058
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-433 <BLAT>
A:CROSS-references: GB:AE000360; GB:U00096; NID:G2367157; PIDN:AACT5808.1; PID:gl789125;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: fixC protein
C:Keywords: oxidoreductase

Query Match 6.9%; Score 160; DB 1; Length 433;
Best Local Similarity 23.3%; Pred. No. 3.9e-05;
Matches 106; Conservative 49; Mismatches 146; Indels 154; Gaps 24;

QY 8 LLIVGAGTGTSTALHLARRGYNVTVDLPYPVPSAISAGNDVNKVISSQYSNKKDEIE 67
DB 17 IIIIGAGIAGTACALRCARAG-LSVLLLERAPIG-----SKNLGGRL----- 59
QY 68 VNEILAEAFNGWKNDPLFKPYVHDITGLMSACSOEGLDRL---GVRVRPGEDPNLVELT 124
DB 60 YTHALAE-----LLPQHFLTAPLERITHESLSLLTPDGYTT----- 96
QY 125 RPEQFRKLAPEG-----VLQDPPGKGYFARSGAGWAHARNALVAAAREAQRMGVKFTVG 180
DB 97 ----FSSLPQGGESWSVLRAFPDW-----LVA---EAEKEGVBCI-- 130
QY 181 TPQGRVVTILFENNVDKGVATGDKIWAERTFLCAGASAGQFLDPKQNLPTATWTLVHI 240
DB 131 -PGATVDALYEENGRVCGVICGD-DILRARYVYVLAEGA-----NSVLAERHGLV-- 177
QY 241 ALKPEERALKNIPVIFN-----IERGFFPEDEERGEI---KICDEHPG-----YTNMVQ 288
DB 178 -TRPAGEAMALGKIVLSLETSIAIERFHLNENGAALIFSGRICDDLPFGAFLYTNQQT 236
QY 289 SADG-----TMMSSIPPEKTIQPKAEATRVALLKET-----MPQ--LADRPFS 329
DB 237 LSLGIVCPSSLTSQSRVPASELUTRFKAHPAVRPLIKNTLESLEYGAHLVPEGGLSHMPVQ 296
QY 330 FARICW-----CADTANREFLDRHPQYHSLVLCGASGRGFKY-----LPSIGNLI 376
DB 297 YAGNGWLLVGDALRSCVNT-----GISVRGMDMALTGQAQAQTL 336
QY 377 VDAMEGKVPQKIHELKWNPDIAANRNWRDITLGRF 411
DB 337 ISACQHREPQNLFPFL--YHNHVERSLIM-DVLQRY 368

RESULT 8
JU0461
sarcosine oxidase (EC 1.5.3.1) - *Bacillus* sp.
C:Species: *Bacillus* sp.
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C:Accession: JU0461
R:Koyama, Y.; Yamamoto-Otake, H.; Suzuki, M.; Nakano, E.
Agric. Biol. Chem. 55, 1259-1263, 1991
A:Title: Cloning and expression of the sarcosine oxidase gene from *Bacillus* sp. NS-129
A:Reference number: JU0461; MUID:91291342; PMID:1368683
A:Accession: JU0461
A:Molecule type: DNA
A:Residues: 1-387 <KQY>
A:CROSS-references: UNIPROT:P23342; GB:D10553; GB:D01067; NID:G216326; PIDN:BAA01410.1;
A:Experimental source: strain NS-129
C:Comment: This enzyme catalyzes the oxidation of sarcosine to generate formaldehyde, g
C:Superfamily: sarcosine oxidase
C:Keywords: oxidoreductase

Query Match 6.7%; Score 155; DB 2; Length 387;
Best Local Similarity 22.8%; Pred. No. 8.5e-05;
Matches 94; Conservative 62; Mismatches 178; Indels 78; Gaps 22;

QY 8 LLIVGAGTGTSTALHLARRGYNVTVDLPYPVPSAISAGNDVNKVISSQYSNKKDEIE 67

Db 7 VIVVGAGSGMAAGYYLAKQGVKTLIV--DSFDPHTNGSHGCDTRIIRHA-YGEGREYVP 64
QY 68 VNEILAEAFNGWKNDPLFKPYVHDITGLIM-----SACSOEGLDRLGVVRVPGEDPNLV 121
Db 65 F-ALPQAQELWYELEKETHHK-IFTQTGVLVYGPKGSAFVSETM-----EAAANH 112
QY 122 ELTRPEQFRKLAPEGVLQDPPGKVG-----YFA-RSGAGWAHARNALVAAAREAQRM 173
Db 113 SLEH-ELF-----EGQLTD--RWAGVEVPDNYEAFIFPNSGVLFSENCIQAYRELAEAH 164
QY 174 GVKFVTGTP--QGRVVVTILFENNVDKGVATGDKIWAERTFLCAGASAGQFLDPKQNL 231
Db 165 GATVITYTPVEDFEVTELDLTIKTAGSYT-----ANKLVVSGAMNSKULSKLD---- 214
QY 232 PTATWTLVHIALKE-----ERALKYN-----IPV-IPNIERGFFFS--PDEERGEIKC 277
Db 215 -----VELPQYQVGVGFECDEAKYSNNAHYAFWVEVENGYIYGFPSGGSLKI- 267
QY 278 DEHPGYTNMVQADGTMMSIPPEKTIQPKAEATRVALLKETMPQLADRPFSFARICWCA 337
Db 268 ----GVHSYGQIIPDTINREFGAY---PEDEANLRKFLQYMPG-ANGELKKGAVCMYT 319
QY 338 DTANREFLDRHPQYHSLVLCGASGRGFKYLPISIGNLIVD-AMSGKVPQKI 388
Db 320 KTPDEHFEVDLHPKYSNVAIAAGFSGHGFKFSVVVGETLAQLATTGKTEHDI 371

RESULT 9
E91081
hypothetical protein Ecs3621 [imported] - *Escherichia coli* (strain O157:H7, substrain RIN
C:Species: *Escherichia coli*
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E91081
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E91081
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <HAY>
A:CROSS-references: UNIPROT:Q8X7T8; GB:BA000007; PIDN:BA37044.1; PID:gl3336092; GSPDB:G
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs3621
C:Superfamily: fixC protein

Query Match 6.4%; Score 150; DB 2; Length 433;
Best Local Similarity 22.9%; Pred. No. 0.00026;
Matches 104; Conservative 49; Mismatches 148; Indels 154; Gaps 24;

QY 8 LLIVGAGTGTSTALHLARRGYNVTVDLPYPVPSAISAGNDVNKVISSQYSNKKDEIE 67
Db 17 IIIIGAGIAGTACALRCARAG-LSVLLLERAPIG-----SKNLGGRL----- 59
QY 68 VNEILAEAFNGWKNDPLFKPYVHDITGLMSACSOEGLDRLGVVRVPGEDPNLVELTRPE 127
Db 60 YTHALAE-----LLPQHFLTAPLERITHESL-----SLTPD 92
QY 128 ---QFRKLAPEG-----VLQDPPGKGYFARSGAGWAHARNALVAAAREAQRMGVKFTVG 180
Db 93 CATTFSSLPQGGESWSVLRAFPDW-----LVA---EAEKEGVBCI-- 130
QY 181 TPQGRVVTILFENNVDKGVATGDKIWAERTFLCAGASAGQFLDPKQNLPTATWTLVHI 240
Db 131 -PGATVDALYEENGRVCGVICGD-DILRARYVYVLAEGA-----NSVLAERHGLV-- 177
QY 241 ALKPEERALKNIPVIFN-----IERGFFPEDEERGEI---KICDEHPG-----YTNMVQ 288
Db 178 -TRPAGEAMALGKIVLSLETSIAIERFHLNENGAALIFSGRICDDLPFGAFLYTNQQT 236

```
QY 289 SADG-----TWMSIPFEKTIQPKAEATVRALLKET-----MPQ--LADRPFS 329
Db 237 LSLGIVCLSSLTOSRVPASELLARFKTHPAVRPLIKNTLSLEYGAHLVPEGGHLSMPVQ 296
QY 330 FARICW-----CADTANREFLIDRHPQYHSLVLCGASGRGPKY-----LPSIGNLI 376
Db 297 YAGNGWLLVGDALRSCVNT-----GISVRGMDTALTGAQAAQTLL 336

QY 377 VDAMEGKVPQKIHELKWNPDIAANRNRWDTLGRF 411
Db 337 ISACQHREPQNLFFL--YHNNVERSLIW--DVLQRY 368

RESULT 10
F85926
hypothetical protein ygcN [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: F85926
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: F85926
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-433 <STO>
A;Cross-references: UNIPROT:Q8X7T8; GB:AE005174; PIDN:AAG57874.1; GSPDB:G
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: ygcN
C;Superfamily: fixC protein

Query Match 6.4%; Score 150; DB 2; Length 433;
Best Local Similarity 22.9%; Pred. No. 0.00026;
Matches 104; Conservative 49; Mismatches 148; Indels 154; Gaps 24;

QY 8 LLIVAGTGTCTSTALHLARRGYTNVTLDPYPVPSAISAGNDVNKVISSQGYNNKDEIE 67
Db 17 IIIIGAGIAGTACALCARAG-LSVLLLEAEIPG-----SKNLGGRL----- 59
QY 68 VNEILAEAEANGKNDPLFPYVYHDTGLMSACSQEGLDLGVVRPGEENLVELTRPE 127
Db 60 YTHALAE-----LLPQFHLTAPLERRITHESL-----SLITPD 92
QY 128 ---QFRKLAPEG---VLQGDFFQWKGYFARSAGWAHARNALVAAAEARQMGVKFVTG 180
Db 93 CATTFSSLPQGGESWSVLRARFDPW-----LVA---EAEKGVVECI-- 130
QY 181 TPQGRVVTLIFENNVDKGAVTGDKIWAERTFLCAGASQGLDFKNQLRPTAWTLVHI 240
Db 131 -PGATVDALVEENRGVCGVCGD-DILRARYVVLAEQA-----NSVLAERHGLV-- 177
QY 241 ALKPEERALLYKNIPVFN-----IERGFFPEDEERGEI---KICDEHPG-----YTNMVQ 288
Db 178 -TRPAGEAMALGIKEVLSLETSIAIEERFHLNENGAALFSGGICDDLPGCAFLYTNQOT 236
QY 289 SADG-----TWMSIPFEKTIQPKAEATVRALLKET-----MPQ--LADRPFS 329
Db 237 LSLGIVCLSSLTOSRVPASELLARFKTHPAVRPLIKNTLSLEYGAHLVPEGGHLSMPVQ 296
QY 330 FARICW-----CADTANREFLIDRHPQYHSLVLCGASGRGPKY-----LPSIGNLI 376
Db 297 YAGNGWLLVGDALRSCVNT-----GISVRGMDTALTGAQAAQTLL 336

QY 377 VDAMEGKVPQKIHELKWNPDIAANRNRWDTLGRF 411
Db 337 ISACQHREPQNLFFL--YHNNVERSLIW--DVLQRY 368

RESULT 11
AB3160
santhopine deaminating protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 16-Aug-2004
C;Accession: AB3160
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AB3160
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-380 <KUR>
A;Cross-references: UNIPROT:Q8UKU5; GB:AE008687; PIDN:AAL45696.1; PID:gl7743424; GSPDB:G
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: agaE
C;Superfamily: Sarcosine oxidase

Query Match 6.4%; Score 148; DB 2; Length 380;
Best Local Similarity 22.2%; Pred. No. 0.00031;
Matches 90; Conservative 51; Mismatches 151; Indels 114; Gaps 18;

QY 6 SLLIVGAGTGTSTALHLARRGYTNVTLDPYPVPSAISAGNDVNKVISSQGYNNKDE 65
Db 12 SSAVIIGAGIFGVSTGQLARRG-IQVTILNDGP-FANGASGRSLWLSARSRSEPHQ 69
QY 66 IEVNEI-----LAEE-----AFNG-----WKND-----PLFKPYVYHDTGLMSACSQEGLD 106
Db 70 LRMAGIDRYRTLAAENPDVEWLRFDGLTWDSDERNEIDAAVYRHEVSLAYDA-----Q 123
QY 107 RLG-----VVRPGEENLVELTRPEQFKLAPGVLTQGDPPGWKGYFARSAGWAHARNA 162
Db 124 RLSAGDVARVTPGIDAGAI-----TPQGA-----FNPGEGMVDLPTL 161
QY 163 LVAAAEARQMGVKFVTGTPQGRVVTLIFENNVDKGAVTGDKIWAERTFLCAGASQ 222
Db 162 IRVLLIEFPALGGVLTQDGAARV---MLEGGRVAGAEATAGVYLRADAVVLTGPAVPK 218
QY 223 FLDFKNQLRPTAWTLVHIALKPEERALLYKNIPVFNIER-----GFFPEPDEERGEIK 275
Db 219 MVGESQIIIGDG---TPIALLVQTKPLAHLPLRAVLNTPRVAVRPAFGSGFSLDADWA--- 272
QY 276 ICDEHFGYTNMQSADGTMSPFEKTIQPKAEATVRALLKETMPQLADRPFSFARICW 335
Db 273 -ADE-----GVTVRADGTY-----EIDDTIVAEILLVEAAK----- 301
QY 336 CADTANREFLIDRHPQYHSLVLCGASGRGFKYLPISGLNLIVDAME 381
Db 302 -----VMEGNPOLEVASIGVGG-----KPIPDGGEPTVGAIK 333

RESULT 12
G83298
conserved hypothetical protein PA2776 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C;Accession: G83298
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brj
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: G83298
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-427 <STO>
A;Cross-references: UNIPROT:Q91067; GB:AE004705; GB:AE004091; NID:g9948851; PIDN:AAG0616
```

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA2776

C;Superfamily: hypothetical protein HI0499

Query Match 6.3%; Score 146.5; DB 2; Length 427;

Best Local Similarity 23.7%; Pred. No. 0.00049;

Matches 98; Conservative 59; Mismatches 164; Indels 93; Gaps 23;

QY 6 SSSLIVGAGTGTSTALHARRGYTNVTLDPYPVPSAISAGNDVNKVISSGYSNNKDE 65

DB 30 TDVCVIGAGTGLSTALFLENGF-KVTVLEAAKVGFGAS-GRNGGQIVNS--YERDIDV 85

QY 66 IEYN-----EILAEAFNGKNDPLFKPKYTHDTGL-----LMSACSQBLDLRLGVR 111

DB 86 IERTVKREAQLGEMAFEGGR---IIRERVARYGICQDLKDGGVFAAFTEKQMDHLRAQ 142

QY 112 VRPGE--DNLVELTRPEQFRKLAPGCVLQGFPGWKGYFARSGAGWAHARNALVAARE 169

DB 143 KQLWERYGHNLQLEIMDAKRIR-----EVVATDNYIG--GMLMSG-GHIHPLNLALGEAAA 195

QY 170 AORMGVKFVTGTPQGRVVTLIFENNNDVGAVTGDGKIWAERTFLCAGASAGQFLDFKNQ 229

DB 196 VESLGRRIIVEQSPATRI-----ERGASPVVHTPOGKV---KAKPIVVAGNA--YLN---- 241

QY 230 LRPTAWTLVHIAKPEERALKYKNIP-----VI-----FNIERGFFFEPEDEERGEIKICDEHP 281

DB 242 -----GLVPELAA--KSMPCGTQVITTEPLNEELAHSLLPQD-----YCVEDC 282

QY 282 GYTNMVQSADGTWMSI-----PPEKTOIPKEAETRVALLKETWMPQLADRPFPARICWCA 337

DB 283 NYLLDYRLSGDKRLIYGGGVYIGARDPADIEAIRPKMLKTFPQKQVKIDFA--W-- 337

QY 338 DFRANREFLDRHPQ-----VHSLVLCGASGRGFKYLPSTGNLIVDAMEGK 383

DB 338 -TGNFLLTSRLPQVGRIGDNIYYIQ-----GCSGHGVYTHVAGKRVIAELRQK 386

RESULT 13

S16133

dimethylglycine dehydrogenase (EC 1.5.99.2) - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004

C;Accession: S16133

R;Lang, H.; Polster, M.; Brandesch, R.

Eur. J. Biochem. 198, 793-799, 1991

A;Title: Rat liver dimethylglycine dehydrogenase. Flavinylolation of the enzyme in hepatoc

A;Reference number: S16133; MUID:91266966; PMID:1710965

A;Molecule type: mRNA

A;Residues: 1-857 <LAN>

A;Cross-references: UNIPROT:Q63342; EMBL:X55995; NID:g56688; PIDN:CAA39468.1; PID:g56688

C;Keywords: flavoprotein; oxidoreductase

Query Match 6.1%; Score 142.5; DB 2; Length 857;

Best Local Similarity 20.3%; Pred. No. 0.0028;

Matches 98; Conservative 74; Mismatches 160; Indels 151; Gaps 23;

QY 9 LIVGAGTGTSTALHARRGYTNVTLDPYPVPSAISAGN-----DVNKV 53

DB 46 VIIGGGCVGSYLAHLAKAGMRDVILLE-----KSELTAGSTWHAAGLTTFYFHPGINLKKI 101

QY 54 ISSGQYSNNKDETEVNEILAEAFN--GWKNDPLFKPYVHDTGLLMSACSQBLDLRLGVR 111

DB 102 -----HYDSIKLVERLEETGQVVG-----FHQPSIRLATTPERVDFEKYQ 143

QY 112 V-RPGEDPNLVELTRPEQFRKLAP-----EGVLQDFFPGWKGYFARSGAGWAHARNALVAA 166

DB 144 MTRTNWHTAQYIIEPKIHELFPULLNMDKILAGY-----NPGDGHIDPYSLTMAL 195

QY 167 AREAORMGV--KF---VTG-----TQGRVVTLIFENNNDVGAVTGDGKIWA 209

DB 196 ATGARKYGVLLKYPAPYTSLSKPRPDGTWQVETPQGSV-----RA 234

QY 210 ERTFLCAG---ASAQFLDFKNQLRPTAWTLVHIAKPEERALKYKNIPVIFNIE----- 260

DB 235 NRIVNAAGFWAREVGVKMIGLDHPLIPVQHQYVVTSTIPEVKALKRELPVLRLDLEGSYYLR 294

QY 261 ---RCGFFPEPDEERGEIKICDE-----HPGYTNMVQSADGTWMSIPEKTOIPKEAETR 311

DB 295 QERDGLLFGPYESQBMKMLQASWVAHGVPPGFGKLFESDLDDRI-----TEHVEAAMEM 348

QY 312 VRALLKETWMPOLADRPFPARICWCADTANREFLDRHPQVHSLVLCGASGRGFKYLP 371

DB 349 VPLKAKLIINIVNGPIIYS-----PDILP---MVGPHQGVNRYVAIG-FGYGIIHAGG 399

QY 372 IGNLIVD-AMEGKVPQKTHELIKMNPDIAANRNWRDITLGRFGGPNRVMDFDHVKWNTNVQ 430

DB 400 VGKYLSDMLIHGEPP---FDLIELDPN-----RYG-----KWTITQ 432

QY 431 YRD 433

DB 433 YTE 435

RESULT 14

B83078

probable D-amino acid oxidase PA4548 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 16-Aug-2004

C;Accession: B83078

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoquchi, S.D.; Warrenner, P.; Hickey, M.J.; Bro

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: B83078

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-364 <STO>

A;Cross-references: UNIPROT:P33642; GB:AE004868; GB:AE004091; NID:g9950785; PIDN:AAG07936

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA4548

C;Superfamily: Sarcosine oxidase

Query Match 6.0%; Score 139.5; DB 2; Length 364;

Best Local Similarity 24.6%; Pred. No. 0.0014;

Matches 103; Conservative 44; Mismatches 165; Indels 107; Gaps 19;

QY 5 SSSLIVGAGTGTSTALHARRGYTNVTLDPYPVPSAISAGNDVNKVISSGQYSNNKD 64

DB 2 SERVVVVGAGVIGLLLTARELALAGLRVTLVERGESGREASWAGGGIVSPLYPMRYS- 58

QY 65 EIEVNEILAEAFNGKNDPLFKPYVHDTGLLMSACSQBLDLRLGVRVPGEDPNL- 120

DB 59 -----AVTALAHWSQD--FYP-----ALGQRLLDETGL-----DPEVHTVG 92

QY 121 ---VEL-----TRPEQFRKLAPGVLQDFFGKWGYFARS--GAGWAHARNA 162

DB 93 LYWLDDDDQTEALQWARKHTRP---LKEVPIEEAAVAPGLGAGFORAVVMGSAVANRNP 149

QY 163 LVAAREAQRMGVKFTVTPQGRVVTLIFENNNDVGAVTGDGKIWAERTFLCAGASAGQ 222

DB 150 RLARSLSRSLQOQFANLELHEQTEVRGWLDRGDRVVGAVTSRGEI-RGDKVLAAAGWSG 208

QY 223 FLDPKNQLRPTAWTLVHIAKPEERALKYK---NIPVIFNIEGPPFFPEDEERGEIKICD 278

DB 209 L-----LKLGLGELPVPVPK-GOMLLYKCAADFLPRMV-LAKGRVAIIPRD-CHILIGS 259

QY 279 --EHPGYTNMVQSADGTWMSIPEKTOIPKEAETRVALLKETMPQLAD-----R 326

DB 260 TLEHSG-----FDKTP-TDEAQESLRASAAELLPELADMOPVAHWAGLR 302

QY 327 PFSFARICWCADTANREFLDRHPQVHSLVLCGASGRGFKYLPSTGNLIVDAMEGKVP 385

Db 303 PGSEGPY-----IGVPGDGLWNTGHYRGLVLAPASCLLLDMSREP 351

RESULT 15

A83905
hypothetical protein BH2041 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 16-Aug-2004
C:Accession: A83905
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Saeaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A83905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <STO>
A:Cross-references: UNIPROT:Q9KB87; GB:AF001514; GB:BA000004; NID:G10174613; PIDN:BA057
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2041
C:Superfamily: Sarcosine oxidase

Query Match	5.9%;	Score 138;	DB 2;	Length 376;
Best Local Similarity	22.3%;	Pred. No. 0.002;		
Matches	90;	Conservative 62;	Mismatches 185;	Indels 66; Gaps 18;

QY	8	LLIIVGAGTGTGTTALHARRGYTNVTYLDPPVPSAISAGN-DYKNVISSQYSNNKDEI	66
DB	7	LCIIGGVIGLSTAIHTAKKG-ANVLVEKDRLCQEASGGAGLLITVAEG-----VDRH	60
QY	67	EYNEILAEAFN---GWNKDPLFKPYHYDHTGLLMSACSQEGDLRLGVYRVRPG---EDPNL	120
DB	61	DPYSIFSRESIQYLLOW-----IPELEEKTOISSGLVKQNLRLALHEQEAALLDEHWL	114
QY	121	VELTRPQFRKLAP-----EGVLQGGPFGWKGYFARSGAGWAHARWALV--AAAREAQ	172
DB	115	MQETNSQIKRLSABEARTLEPLSNDL-----IMACSPDEYHLEPSLYLNALARWQ	169
QY	173	MGVKFVTGTQGRVVTULIFENNDDYKGAVTGDGKIWRAERTFLCAGASA--GQFLDFKNQ	229
DB	170	LG---VTIREQTEVIQFIHQNNRMVGVVNTGTQIVAEETTLAAGANSHLLSLRLAKVKVP	226
QY	230	LRPTAWTLVHIALKPEERALYKNIPLVFENTERGFFPDEBERGIIKIDHEHPGVTNNVQS	289
DB	227	VYPLKGQL---IVLKGQD-----TDLVNVMINTSKGYILSKPNQTVTVLGGATSEKEKNKI-T	279
QY	290	ADGTMMSGIP-PEKTOIKPEARETRVALLKETMPQLADRPFSFARICWCADTANREFILDR	348
DB	280	AKGIQELLPLFD--YVPELKWWEIQHLWAGLRPA-----TPDQOPIIGY	321
QY	349	HPQVHSVLVCGASGRGFKYLPISIGNLIIVDAMEGKVPQKITHEL	391
DB	322	VPEWEGWYLATGHSRRHGVLSSGWTGHLVAABELEGK--DQSHKL	362

Search completed: May 27, 2005, 13:48:43
Job time : 35.5437 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:38:36 ; Search time 51.2473 Seconds
(without alignments)
975.171 Million cell updates/sec

Title: US-10-622-893A-3

Perfect score: 2329

Sequence: 1 AVTKSSSLIVGAGTGTST.....MDPHDVKEWTVQYRDISKL 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 538906 seqs, 114359116 residues

Total number of hits satisfying chosen parameters: 538906

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New:

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pap.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pap.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pap.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pap.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pap.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pap.*
- 7: /cgn2_6/ptodata/2/paa/US11_NEW_COMB.pap.*
- 8: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	215	9.2	459	8	US-60-643-717-13558
2	181.5	7.8	440	7	US-11-097-143-4818
3	177.5	7.6	390	6	US-10-990-328A-13084
4	177.5	7.6	445	6	US-10-990-328A-13083
5	169	7.3	387	6	US-10-990-477-6
6	167	7.2	389	6	US-10-534-583-1
7	165	7.1	387	6	US-10-990-477-2
8	148.5	6.4	774	6	US-10-990-328A-13788
9	148.5	6.4	866	6	US-10-990-328A-13787
10	142	6.1	388	6	US-10-526-324-372
11	130.5	5.6	386	6	US-10-526-324-29
12	124.5	5.3	380	8	US-60-669-241-28797
13	115.5	5.0	389	1	PCT-IB03-06509-1503
14	115	4.9	505	7	US-11-031-175-9986
15	108	4.6	796	6	US-10-450-763-48529
16	108	4.6	879	6	US-10-204-639-58
17	106.5	4.6	822	8	US-60-643-717-17883
18	105.5	4.5	822	8	US-60-643-717-7212
19	104.5	4.5	339	7	US-11-122-943-10
20	104.5	4.5	339	7	US-11-122-943-12
21	104	4.5	515	7	US-11-097-143-42765
22	100.5	4.3	102	8	US-60-655-875-133298
23	99.5	4.3	340	7	US-11-122-943-18
24	98.5	4.2	849	8	US-60-643-717-18183
25	98.5	4.2	850	8	US-60-643-717-17113

ALIGNMENTS

RESULT 1

US-60-643-717-13558
; Sequence 13558, Application US/60643717
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53629)A
; CURRENT APPLICATION NUMBER: US/60/643,717
; CURRENT FILING DATE: 2005-01-12
; NUMBER OF SEQ ID NOS: 19247
; SEQ ID NO 13558
; LENGTH: 459
; TYPE: PRT
; ORGANISM: ASPERGILLUS NIDULANS FGSC A4
US-60-643-717-13558

Query Match 9.2%; Score 215; DB 8; Length 459;
Best Local Similarity 19.5%; Pred. No. 1.5e-10;
Matches 82; Conservative 34; Mismatches 84; Indels 220; Gaps 10;

Qy	3	TKSSLLIVGAGTGTSTALHARGYNTVTLD-PYPVPSAISAGDNVKNKVISSGQYSN	61
Db	5	SKSIPTAIIIGDAGFLSTALHLVONGYDTITVLEQDEKIPPPYSAANYLNKIV-----	57
Qy	62	NKDEIEVNEILABEAFNGWKNDFLPKPYHDTGLLMSACSOEGLDRLGVVRPGEEDNLV	121
Db	58	-REY-----EDP----	64
Qy	122	ELTRPEQFKLAPEGLVQDFPGWKGYFARGAGAHARNALVAAAREAQRMGVKFTGT	181
Db	65	-----	64
Qy	182	PQGRVVTLIPENNVDKVGATGDKIWAERTFLCAGASAGQFLDFKNQLRPTATWLHIA	241
Db	65	-----	64
Qy	242	LKPEERALKYNIPIVFIPIERGFPEPDEERGEIKICDEHPGYTMVOSADGTWMSIPFEK	301
Db	65	-----LCPMGSGYIN-TDKATGVSHS-PFPV	95
Qy	302	TPKEATRVALLKETMQLADRPSPARICWCADTANREFLIDRHP-QYHS	354
Db	96	NOEASGSLPAEDETIRKLQOTLPALANRPLVLKSLCWFADTKDSDFIIDFVPGSKGS	155
Qy	355	LVLGCGASGRGFKYLPISGLNLIVDAMEGKVPQKIHILIKWNPDIANRNWDTLGRGGP	414
Db	156	VTIESADSGHGFQWFFIVGWSLTSKRTPTNSIH-----SKPNLRANMGQSTVGNVGLP	212


```

Best Local Similarity   21.7%; Pred. No. 3.1e-07;
Matches      86; Conservative    66; Mismatches 174; Indels     71; Gaps    16;

Qy      9 LIVGAGTGTSTALHLARRGYNTVTLDPYPVPSAIGAGNDVNKVISSGOYSNNKDIEV  68
       :::|||:|||:|||:  :::::  :::::  :::::  :::::  :::::  :::::
Db      66 IVIGAGTCCTAYHLAGH-RKRILLLEOFPFLPHSRGSSHQSRIIRKAYLE-----DF 118

Qy      69 NEILAEAFNGWKNDPLFKPYHYHTDGLLMSACSEGLDRIGVRVRPGEDPNLVLTPEQ 128
       :::|||:|||:|||:  :::::  :::::  :::::  :::::  :::::  :::::
Db      119 YTRMHCEYCIWAQ-----LEHEAGTQLH--RQTGLLLGMKENQELKTITQANLSQRV 170

Qy      129 FRKLAPGVLGDDP-----GWKGYPARSGAGWAHARNALVA---AARE---AORMGVK 176
       :::|||:|||:|||:  :::::  :::::  :::::  :::::  :::::  :::::
Db      171 EHOCISSELKORFENIRLPKEVGELLDNSG-GVIYAKALRALQDAIROLGGIVRDGEK 229

Qy      177 FVTGTPQGRVVTTLIFENNVDKGAVTGDGKIWRAERTFLCAGASAGQFLDKFNQLRPATAWT 236
       :::|||:|||:|||:  :::::  :::::  :::::  :::::  :::::  :::::
Db      230 VVEINP-GLAVT-----VKTTSRSQAOKSLVITAGPTWNQL-----LRPLGIE 271

Qy      237 LVHIALKPEERALKYNIPVFINIERGP-----FFPDDEERGEIKTCDHPHG 282
       :::|||:|||:|||:  :::::  :::::  :::::  :::::  :::::  :::::
Db      272 MPLQTLRLNVCYWRMVPGSYGVSQAFCFLWLGLCPHHIYGLTGTVPGMLMKVSYHHGN 331

Qy      283 YTNMVQSADGTMMSIPEKTIQPKAEATRV-RALLKETMPOLADRPSPFARI CWCADTAN 341
       :::|||:|||:|||:  :::::  :::::  :::::  :::::  :::::  :::::
Db      332 H-----ADPERDCPTARTDI---GDVQILSSFVRDLPLDKPEP-AVIESCMYTNTPD 381

Qy      342 REFLIDRHPOVHSVLVCGSAGRKFYKPLSIGNLIVD 378
       :::|||:|||:|||:  :::::  :::::  :::::  :::::  :::::  :::::
Db      382 EQFILDRHPKYDNIVIGAGFSGHGFKLAPVVGKILYE 418


RESULT 5
US-10-990-477-6
; Sequence 6, Application US/10990477
; GENERAL INFORMATION:
; APPLICANT: KIRKOMAN CORPORATION
; APPLICANT: FURUKAWA, Keisuke
; APPLICANT: KAJIYAMA, Naoki
; TITLE OF INVENTION: MODIFIED SARCOSINE OXIDASES, MODIFIED SARCOSINE OXIDASE GENES,
; TITLE OF INVENTION: AND METHODS FOR PREPARING THE MODIFIED SARCOSINE OXIDASES
; FILE REFERENCE: 261714USO
; CURRENT APPLICATION NUMBER: US/10/990,477
; CURRENT FILING DATE: 2004-11-18
; PRIOR APPLICATION NUMBER: JP 2003-387975
; PRIOR FILING DATE: 2003-11-18
; PRIOR APPLICATION NUMBER: JP 2004-184960
; PRIOR FILING DATE: 2004-06-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Bacillus sp. strain KS-11a
US-10-990-477-6

Query Match          7.3%; Score 169; DB 6; Length 387;
Best Local Similarity 21.4%; Pred. No. 1.5e-06;
Matches      89; Conservative    69; Mismatches 172; Indels     86; Gaps    20;

Qy      8 LLIVGAGTGTSTALHLARRGYNTVTLDPYPVPSAIGAGNDVNKVISSGOYSNNKDIEI  67
       :::|||:|||:|||:  :::::  :::::  :::::  :::::  :::::  :::::
Db      7 VIIVGAGSMGMAAGYLAKQGVKTLV-DAFDPHTEGSHHGDTRIIRHA-YGBGREVVP  64

Qy      68 VNEILASEAFNGWKNDPLFKPYHYHTDTGLL-----MSCAQSQL-----D 106
       :::|||:|||:|||:  :::::  :::::  :::::  :::::  :::::  :::::
Db      65 F-ALRAQELMWYLELENETHNK-IFTKTGVLVPGKGBSDFVAETMEAAAHSLTVDLLEGD 122

Qy      107 RLGVVRVPGEDPNLVLTREOPKLAPEGVLQDFFGWKGYFARSAGWAHARNALVAA 166
       :::|||:|||:|||:  :::::  :::::  :::::  :::::  :::::  :::::
Db      123 EINTR-WFG-----ITVPENYAI-----PEPNSVLFSNCIRS 157

Qy      167 AREAORMGVKFVTGTPQGRVVTTLIFENNVDKGAVTGDGKIWRAERTFLCAGASAGQFLDF 226

```

```
Db 305 MPG-ATGELKSGAVCYMTKTPDEHFVLDLHPQFSNVAIAAGFSGHGKFFSVVGETLSQL 363
QY 379 AMEGKVPOKI 388
Db 364 AVTGKTEHDI 373

RESULT 7
US-10-990-477-2
; Sequence 2, Application US/10990477
; GENERAL INFORMATION:
; APPLICANT: KIKKOMAN CORPORATION
; APPLICANT: FURUKAWA, Keisuke
; APPLICANT: KAJIYAMA, Naoki
; TITLE OF INVENTION: MODIFIED SARCOSINE OXIDASES, MODIFIED SARCOSINE OXIDASE GENES,
; TITLE OF INVENTION: AND METHODS FOR PREPARING THE MODIFIED SARCOSINE OXIDASES
; FILE REFERENCE: 261714US0
; CURRENT APPLICATION NUMBER: US/10/990,477
; CURRENT FILING DATE: 2004-11-18
; PRIOR APPLICATION NUMBER: JP 2003-387975
; PRIOR FILING DATE: 2003-11-18
; PRIOR APPLICATION NUMBER: JP 2004-184960
; PRIOR FILING DATE: 2004-06-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Construct
US-10-990-477-2
Query Match 7.1%; Score 165; DB 6; Length 387;
Best Local Similarity 21.4%; Pred. No. 3.3e-06;
Matches 89; Conservative 68; Mismatches 173; Indels 86; Gaps 20;
```

```
QY 8 LLIVGAGTGTSTALHLAGRYNTVTLDPYPVPSAISAGNDVKNVYISSGOYSNNKDEIE 67
Db 7 VIVVGAGSMGMAAGYYLAKQGVKTLIV-DAPDPHTEGSHHGDTIRIHA-YGEGREYVP 64
QY 68 VNEILAEAFNGMKNDFLEKPYHYDHTGLL-----MSACSOEGL-----D 106
Db 65 F-ALRAQELWYLENEHTNK-IFTKTGLVFGPKGSDFAETMEAAAEHSLIVDLEGD 122
QY 107 RLGVVRPGEDPNLVELTRPEQPRKLAPEGVLQGDPPGKGYFARSGAGWAHARNALVAA 166
Db 123 EINTR-WPG-----ITVPENYNAI-----FEPNSGVLPFSENCIRSY 157
QY 167 AREAQRMGVKTGTGTPQGRVVTLLIFENNVDKGNVTDGDKIWAERTFLCAGASAGQFLDF 226
Db 158 RELAVAKGAKILTYT---RVEDFEVSQDQVK-IQTANGS-YTADKLIVSGAWNSKLSK 212
QY 227 KNQLRPTAWTLVHIALKP-----EERALKN---IPV-IFNTERGFFPEPDEERGE 273
Db 213 LN-----LDIPQPRQVGVGFDSEAKYSNDVDYPAFVVEVPKGIYYGFPFSGG- 262
QY 274 IKICDHPGYTNMVSADGTMSIPPEKTQIPKEAETRVALLKETMPLADRPFSFARI 333
Db 263 ---CGLKIGVHTYQQQIDPDTINREFGAYQ---EDESNLRDFLEKYMPE-ANGELKRGAA 315
QY 334 CWCADTANREFLDIRHPQVHSLVLCGASGRGFKYLPSTGNLIVD-AMEGKVPOKI 388
Db 316 CMYTKTPDEHFVIDTHPEHSNVEVAAGFSGHGKFFSVVGEVLSQATTGKTEHDI 371
```

```
RESULT 8
US-10-990-328A-13788
; Sequence 13788, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
```

```
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328A
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13788
; LENGTH: 774
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-990-328A-13788
Query Match 6.4%; Score 148.5; DB 6; Length 774;
Best Local Similarity 20.9%; Pred. No. 0.0025;
Matches 98; Conservative 73; Mismatches 176; Indels 121; Gaps 22;
```

```
QY 9 LIVGAGTGTSTALHLAGRYNTVTLDPYPVPSAISAGN-----DVNKV 53
Db 53 VLIIGCGVGVSLAVHLAKAGMKDVVLE-----KSELTAGSTWHAAGLTYTFHPGINKKI 108
QY 54 ISSGOYSNNKDEIEVNEILAEAFN--GWKNDPLPKPYHYDHTGLMSACSQGLDRLGVR 111
Db 109 -----HYDSIKLYEKLLEETGQVVG-----FHQPGSIRLATTPVRVDEFKYQ 150
QY 112 V-RPGEDPNLVELTRPEQPRKLAPE-----EGVLQGDPPGKGYFARSGAGWAHARNALVAA 166
Db 151 WRTGWHATEQYLIIEPEKIQEMFPLNNMKVLAGLY-----NPGDGHIDPYSUTMAL 202
QY 167 AREAQRMGVKTGTGTPQGRVVTLLIFENNVDKGNVTDGDKIWAERTFLCAG---ASAGQF 223
Db 203 AAGARKCGALLKVPAP---VTSLKARSDGTWDETPOGSM-RANRIVNAAAGFAREWVGKM 258
QY 224 LDFKNQLRPTAWTLVHIALKPEERALKYKNIPIVFNIE-----RGFFPEPDEERGEI 274
Db 259 IGLEHPLIPVQHYVVTSTIPEVKALKRELPLVRLDELSYILRQERDGLLFGYESQEKM 318
QY 275 KICDE-----HPGYTNMV--QSDGTMSIPPEKTQIPKEAETRVALLKETMPLADR 326
Db 319 KVQDSWVTNGVPPGKELFESDLRIMEHI-----KAAMWVPLKADIINVNG 370
QY 327 PFSFARICWADTANREFLDIRHPQVHSLVLCGASGRGFKYLPSTGNLIVD-AMEGKV 385
Db 371 PITYS-----PDILP---MVGPHQGVRYNVAIG-FGYGIIHAGGVKYLSDWILHGEPP 421
QY 386 QKIHELKKNPDIANRNWRDILGRFGGNRVMDPHDKVKEWTNVQYRD 433
Db 422 ---FDLIELDPN-----RYG-----KWTTOYTE 442
```

```
RESULT 9
US-10-990-328A-13787
; Sequence 13787, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328A
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13787
; LENGTH: 866
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-990-328A-13787
Query Match 6.4%; Score 148.5; DB 6; Length 866;
Best Local Similarity 20.9%; Pred. No. 0.0003;
Matches 98; Conservative 73; Mismatches 176; Indels 121; Gaps 22;
```


Db 360 DGKTDK 365

RESULT 11

US-10-526-324-29

Sequence 29, Application US/10526324

GENERAL INFORMATION:

APPLICANT: Imanaka, Takayuki

TITLE OF INVENTION: METHOD OF TARGETED GENE DISRUPTION, GENOME OF

TITLE OF INVENTION: HYPERTHERMOSTABLE BACTERIUM AND GENOME CHIP USING

FILE REFERENCE: 490051.40IUSPC

CURRENT APPLICATION NUMBER: US/10/526,324

PRIOR FILING DATE: 2005-02-28

PRIOR FILING DATE: 2003-08-29

PRIOR FILING DATE: 2002-08-30

NUMBER OF SEQ ID NOS: 2167

SOFTWARE: PatentIn version 3.1

SEQ ID NO 29

LENGTH: 386

TYPE: PRT

ORGANISM: Thermococcus kodakaraensis KOD1

FEATURE:

NAME/KEY: misc feature

LOCATION: (414542)..(414542)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (786890)..(786890)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (786907)..(786907)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (786944)..(786946)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (839139)..(839139)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1128488)..(1128488)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1128499)..(1128499)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1128505)..(1128506)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1128517)..(1128518)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1128539)..(1128540)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1561400)..(1561400)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1561477)..(1561477)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1561545)..(1561545)

OTHER INFORMATION: n is a or c or g or t.

FEATURE:

NAME/KEY: misc feature

LOCATION: (1767941)..(1767941)

OTHER INFORMATION: n is a or c or g or t.

US-10-526-324-29

Query Match 5.6%; Score 130.5; DB 6; Length 386;

Best Local Similarity 21.0%; Pred. No. 0.0037;

Matches 93; Conservative 65; Mismatches 164; Indels 121; Gaps 23;

QY 4 KSSSLIVGAGTGWGTSTALHLARRGYTNVVLDPVPSPSAISAGNDVNKVISSGQYSNNK 63

Db 8 EKSEITIIGGGIVGVTTIAHELAKRG-EVTVIE-----KRFIGSGSTFRCGTGIRQOFNDE 62

QY 64 DEIEVNEILAEAFNGWKNDPLFKPY-----YHDTGLILMSACSQEGLD----- 106

Db 63 ANVQWKRSE-----LWKYSEYGFPPQGTGYLFLLYDDEEVEVTEKRNIAIQN 112

QY 107 RLGVVRPGEDNLVELTRPEQFRKLAP-EGVLQGDFFGKGYFARSGAGNAHARNALVA 165

Db 113 KFGVPTR-----LITPEAKEIVPLLDISEVVAASWNPDTGKASPFHSTAKFAL-- 161

QY 166 AAREAQRMGVKPVGTGTPQGRVVTLIFENNVDKGVATGDKIWRABRTFLCAGASAGQFLD 225

Db 162 ---HAEEFGAKLVEYT---EVKDFIENGIEIKGLKTSRGTI-----KTGIVVNATNA----- 207

QY 226 FKNQLRPTAWTLVHIALKPEERALKYNIPIVNIERNERGFPEDEERGEIK--ICDEHPGY 283

Db 208 WAKLINAMAGITKPIEP-----YKHQAVI-----TOPIKKGSVKPWISFRYGH 253

QY 284 TMMVQSADGTM-SIPFEKTQIPKBAETRVALLKETWPLQADRRFSPFARICWCADTANR 342

Db 254 AVLTQTSHGGIIGVGVE-----EGPTYD---LNPVTEFLREVSYFTKII---PALR 300

QY 343 BELI-----DRHP-----QYHSLVLCGAGSGPGKYLPSIGNLIVDAM-EGK 383

Db 301 ELLILRTWAGYYAKTPDSNPAIGKIELSDDYYIAAGFSGHGPMAPAVAEVADLITKGR 360

QY 384 VPQKIHELKKNPDIAANRNWRD 406

Db 361 T-----DLPW-----WYD 369

RESULT 12

US-60-669-241-28797

Sequence 28797, Application US/60669241

GENERAL INFORMATION:

APPLICANT: Monsanto Technology LLC

APPLICANT: Baum, James A

APPLICANT: Gilbertson, Larry A

APPLICANT: Kovalic, David K

APPLICANT: LaRosa, Thomas J

APPLICANT: Lu, Maolong

APPLICANT: Munyikwa, Tichifa R. I.

APPLICANT: Roberts, James K

APPLICANT: Wu, Wei

APPLICANT: Zhang, Bei

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CONTROL OF INSECT INFESTATION IN PLAN

FILE REFERENCE: 38-21(53596)

CURRENT APPLICATION NUMBER: US/60/669,241

CURRENT FILING DATE: 2005-04-07

PRIOR APPLICATION NUMBER: 60560842

PRIOR FILING DATE: 2004-04-09

PRIOR APPLICATION NUMBER: 60565632

PRIOR FILING DATE: 2004-04-27

PRIOR APPLICATION NUMBER: 60579062

PRIOR FILING DATE: 2004-06-11

PRIOR APPLICATION NUMBER: 60603421

```
; PRIOR FILING DATE: 2002-03-06
; NUMBER OF SEQ ID NOS: 5809
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1503
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Mycobacterium paratuberculosis
PCT-IB03-06509-1503

Query Match      5.0%; Score 115.5; DB 1; Length 389;
Best Local Similarity 22.2%; Pred. No. 0.081;
Matches 53; Conservative 41; Mismatches 96; Indels 49; Gaps 24;

QY      8 LLIVGAGTGTSTALHLARGYNVTVDPPVPVPSAISAGNDVNKVISSQYSNNKDEIE 67
       :::|||   |   |:|:|:|   |||   |   |   |   |   |   |   |   |
Db     74 IIVIGAGVCGLAAAHSLSRGHDDVVVLEKGQPGFGQSAG-----LARIPIAHRRESLC 128

QY      68 VNEILAEAFNGWKNDPLPKFPYYHDTGLMSACSQSG-----LDRLGVVRVPPGEDPNLVEL 123
       :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db    129 RLAFARAGAQWRWEAFVGVRLLGSFGFTAAAGAAADGVAQAMERAGA----- 175

QY    124 TRPEQRFKLAPEGVLQ----GDPFGMKGYFARSAGWAHAARNALVAAAQAORMGVKFVT 179
       :   :   :   :   :   :   :   :   :   :   :   :   |   |   |   :
Db    176 -----AFSRLDRDGIARIPADVPMWTGTGFDPLG-GSLRIRRALTLIAR---RVVIR--- 224

QY    180 GTPOGRVVTLIFFENNDVKAVGTGDGKIWRAERTFLCAGASAGOF-----LDFKNQLRPT 233
       :   :   :   :   :   :   :   :   :   :   :   :   |   |   |   :
Db    225 ---RGVSVSA---DNGSTVLADGTVLRADRVLICAGVATPKLGPLGVDFVPHTRET 276

RESULT 14
US-11-031-175-9986
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/11/031,175
; CURRENT FILING DATE: 2005-01-08
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9986
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-11-031-175-9986

Query Match      4.9%; Score 115; DB 7; Length 505;
Best Local Similarity 21.9%; Pred. No. 0.13;
Matches 92; Conservative 54; Mismatches 148; Indels 126; Gaps 24;

QY      2 VTKSSLLIVGAGTGTSTALHLARGYNVTVDPPVPVPSAISAGNDVNKVISSQYS- 60
       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db    115 VPAAKVVIIIGGINGLALAYNLARAGETDVVLERGYLCAGASGRGGGVRMWGTPSL 174

QY      61 --NKKDEIEVNEILABE-AFNGW-----KNDLPFKPYHYDHTGLMSACSQEGDLR 107
       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db    175 VELAKRSIELMKGFARELGINVWLROGGVIFLAKTAPVAQRLEINVSL-----HNR 225

QY    108 LGVVRVPPGEDPNLVLTREQRFKLAP-----EGVLQGDPFGWKGYFARS 152
       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db    226 FGVPTR-----LITDEARGIVPGJTMKDCLIASYNPEDGVI---FP-WPFL----- 268

QY    153 GAGWAHARNALVAAAQAORMGVKFYTGTGPQGRVVTLIFFENNNDVKAVGTGDGKIWRAERT 212
       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db    269 ---WGYAQG-----CCKRGVRYETYL-----DVTGFESGGQVRKVKT 304

QY    213 -----FLCAGA---SAGOFLDFKNQLRPNTAWTLVHIALKPEERALYKNIPIFNIE 260
       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
```

Db 305 RGDIACTVTLAAGAWSPQAKLADVLPNEPHR-----HEILSTELKDFLG-PLVSVLD 359
QY 261 RGFFFEPEDEERGEI--KICD-EHPGTYTNMVQSGADGTMMSIPFEKTOIPKEAETRVRAALIK 317
Db 360 SGLYFS-QSMRGEIVGGMGDAKEPAGLNM-----GSTLRFV-----SREAQALM 402
QY 318 ETWPLQADRPFFARIC--W--CAD-TANREFLIDRHPOYHSLVLGCGASGRGFKYLPISI 372
Db 403 EQLPEV-----GHVKVLRQWAGCYDVTDPNNPILGRTPLGLDNLQMSGFVGHGFMMAPAV 457

RESULT 15
US-10-450-763-48529
; Sequence 48529, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 48529
; LENGTH: 796
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (413)..(443)
; OTHER INFORMATION: 2Fe-2S ferredoxins iron-sulfur binding region proteins.
; OTHER INFORMATION: domain identified by eMATRIX, accession number BL00197A, p-value=
; OTHER INFORMATION: 8.342e-14, raw score of 18.23
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (18)..(113)
; OTHER INFORMATION: ParB-like nuclease domain identified by Pfam, accession name
; OTHER INFORMATION: ParBc, E-value=4.8e-20, Pfam score of 80.0
US-10-450-763-48529

Query Match 4.6%; Score 108; DB 6; Length 796;
Best Local Similarity 20.0%; Pred.No.1;
Matches 87; Conservative 51; Mismatches 118; Indels 178; Gaps 21;
QY 11 VGAGTWTGTALHLARRGYTNVTLPDPYPVPSAISAGDNDVNVKVISSGOYSNNKDEIEV-- 68
Db 77 IGKG-----SSSLKRLKGLVPTCLE-----GTRNQRIATIRHNRARGHQITA 122
QY 69 -NEILAEAFNGWKNDPLPKPYHYDTHGLMSACSQGLDRL----- 108
Db 123 MSEIVRELSQLGWDNDKIGKELGMDSEVLRLKQINGLQELFADROYRSRAWTLKLAQRSA 182
QY 109 -----GVRVR-----PCEDPNLYELTPEQF 129
Db 183 AASNVDSEFLANRDESYVKGSAWNTDSGIARRAPASATGKTHPGPVMVMVNMVMDQAF 242
QY 130 RKLAP--EGVLQDFFPGWKGYFAR-----SGAG-----WAHARNALVAAREAQ 171
Db 243 KKLNPHEHPVLHSD-QGWQYRMYRYQNIILKEHCGGVTPIMSMRWL-AKNRPQADV--- 297
QY 172 RMGVKFTVTGPQGRV-----VTLIFFENDVKGAVTGDGKIWR-----AERTF 213
Db 298 ---VIYNVTPQDVIPEADWRNYPVTLVAENNVTEGFIA--GRLTRELLAGVPDLASRTV 352
QY 214 LCAGASAGQFLPKFKNQLRPTAWTLVHLAKPBERALYKNIPIVFIENIERGF---FPEPDEE 270
Db 353 MTCGPA--PYMDWVEQ-----EVKAL-----GVTRFFKEKFFTP--- 384

QY 271 RGEIKICDEHPGYTNMVQSGADGTMMSIPFEKTOIPKE-----AETRVALLKETWLPOLAD 325
Db 385 -----VAEAATGSLKFTKIQPAREFYAPVGTTLLEAESNNVDDWA- 425
QY 326 RPFSPARICWCADT 339
Db 426 ---ACRAGVCGCCKT 437
Search completed: May 27, 2005, 13:51:55
Job time : 53.2473 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:47:36 ; Search time 4.92964 Seconds
(without alignments)
1189.648 Million cell updates/sec

Title: US-10-622-893A-4

Perfect score: 90

Sequence: 1 KGELEGLPIPNPLRTG 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1462099 seqs, 344972447 residues

Total number of hits satisfying chosen parameters: 1462099

Minimum DB seq length: 0

Maximum Match 0%

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	100.0	17	17 US-10-622-893A-4	Sequence 4, Appli
2	90	100.0	472	17 US-10-622-893A-5	Sequence 5, Appli
3	59	65.6	146	9 US-09-923-304-9	Sequence 9, Appli
4	58.5	65.0	326	16 US-10-865-978-37	Sequence 17, Appl
5	58.5	65.0	338	16 US-10-467-243-26	Sequence 26, Appl
6	58.5	65.0	349	16 US-10-467-243-8	Sequence 8, Appli
7	58.5	65.0	368	14 US-10-270-223-4	Sequence 4, Appli
8	58.5	65.0	473	15 US-10-239-652A-6	Sequence 6, Appli
9	58.5	65.0	473	15 US-10-239-652A-8	Sequence 8, Appli
10	58.5	65.0	473	15 US-10-415-232-4	Sequence 4, Appli
11	58.5	65.0	669	9 US-09-213-888-25	Sequence 25, Appl
12	58.5	65.0	669	9 US-09-328-877A-25	Sequence 25, Appl
13	58.5	65.0	669	16 US-10-653-497-25	Sequence 25, Appl

14	58.5	65.0	669	16	US-10-653-517-25	Sequence 25, Appl
15	58.5	65.0	669	16	US-10-653-496A-25	Sequence 25, Appl
16	58.5	65.0	669	17	US-10-653-676A-25	Sequence 25, Appl
17	58.5	65.0	991	16	US-10-865-978-34	Sequence 34, Appl
18	56.5	62.8	253	16	US-10-624-909-20	Sequence 20, Appl
19	56.5	62.8	361	14	US-10-006-780-8	Sequence 8, Appli
20	56.5	62.8	383	9	US-09-883-096-5	Sequence 5, Appli
21	55.5	61.7	570	17	US-10-800-350-386	Sequence 386, App
22	51.5	57.2	135	16	US-10-624-909-192	Sequence 192, App
23	51.5	57.2	181	16	US-10-624-909-86	Sequence 86, Appl
24	51.5	57.2	196	16	US-10-624-909-160	Sequence 160, App
25	51.5	57.2	200	16	US-10-624-909-48	Sequence 48, Appl
26	51.5	57.2	212	16	US-10-624-909-68	Sequence 68, Appl
27	51.5	57.2	233	16	US-10-624-909-172	Sequence 172, App
28	51.5	57.2	236	16	US-10-624-909-56	Sequence 56, Appl
29	51.5	57.2	237	16	US-10-624-909-66	Sequence 66, Appl
30	51.5	57.2	237	16	US-10-624-909-102	Sequence 102, App
31	51.5	57.2	237	16	US-10-624-909-124	Sequence 124, App
32	51.5	57.2	237	16	US-10-624-909-126	Sequence 126, App
33	51.5	57.2	237	16	US-10-624-909-188	Sequence 188, App
34	51.5	57.2	238	16	US-10-624-909-52	Sequence 52, Appl
35	51.5	57.2	238	16	US-10-624-909-78	Sequence 78, Appl
36	51.5	57.2	238	16	US-10-624-909-84	Sequence 84, Appl
37	51.5	57.2	238	16	US-10-624-909-88	Sequence 88, Appl
38	51.5	57.2	238	16	US-10-624-909-104	Sequence 104, App
39	51.5	57.2	238	16	US-10-624-909-118	Sequence 118, App
40	51.5	57.2	238	16	US-10-624-909-132	Sequence 132, App
41	51.5	57.2	238	16	US-10-624-909-146	Sequence 146, App
42	51.5	57.2	239	16	US-10-624-909-38	Sequence 38, Appl
43	51.5	57.2	239	16	US-10-624-909-44	Sequence 44, Appl
44	51.5	57.2	239	16	US-10-624-909-60	Sequence 60, Appl
45	51.5	57.2	239	16	US-10-624-909-62	Sequence 62, Appl

ALIGNMENTS

RESULT 1

US-10-622-893A-4
; Sequence 4, Application US/10622893A
; Publication No. US20050014935A1
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Sheng
; APPLICANT: Datta, Abhijit
; APPLICANT: Wang, Yuping
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DETERMINATION OF GLYCATED PROTEINS
; FILE REFERENCE: 46692001300
; CURRENT APPLICATION NUMBER: US/10/622, 893A
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 40%-100% identity of the second bacterial leader sequence
US-10-622-893A-4

Query Match 100.0%; Score 90; DB 17; Length 17;
Best Local Similarity 100.0%; Pred. No. 4.7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KGELEGLPIPNPLRTG 17

Db 1 KGELEGLPIPNPLRTG 17

RESULT 2

US-10-622-893A-5
; Sequence 5, Application US/10622893A
; Publication No. US20050014935A1

; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Sheng
; APPLICANT: Datta, Abhijit
; APPLICANT: Wang, Yuping
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; FILE REFERENCE: DETERMINATION OF GLYCATED PROTEINS
; CURRENT APPLICATION NUMBER: US/10/622,893A
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric protein
US-10-622-893A-5

Query Match 100.0%; Score 90; DB 17; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KGEGLPIPNPLLRG 17
|||||
Db 450 KGEGLPIPNPLLRG 466

RESULT 3
US-09-923-304-9
; Sequence 9, Application US/09923304
; Patent No. US20020081612A1
; GENERAL INFORMATION:
; APPLICANT: KATZ, RUTH
; APPLICANT: JIANG, FENG
; TITLE OF INVENTION: DETECTION AND DIAGNOSIS OF SMOKING RELATED CANCERS
; FILE REFERENCE: UTSC:658US
; CURRENT APPLICATION NUMBER: US/09/923,304
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-923-304-9

Query Match 65.6%; Score 59; DB 9; Length 146;
Best Local Similarity 59.3%; Pred. No. 0.32;
Matches 16; Conservative 0; Mismatches 1; Indels 10; Gaps 2;

QY 1 KGEGLPIPNPLLRG 17
|||||
Db 114 KGEGLPIPNPLLRG 140

RESULT 4
US-10-865-978-37
; Sequence 37, Application US/10865978
; Publication No. US20050003416A1
; GENERAL INFORMATION:
; APPLICANT: Berlex Biosciences
; APPLICANT: Wu, Qingyu
; TITLE OF INVENTION: Novel Modified Corin Molecules Having Substitute Activation Sequ
; FILE REFERENCE: 53103AUST1
; CURRENT APPLICATION NUMBER: US/10/865,978
; CURRENT FILING DATE: 2004-06-09
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 37
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: IgK Secretion Signal Sequence, Corin AA 787 to 796, Enterokinase
; OTHER INFORMATION: Site, Corin AA 802 to 1042, and C-Terminal V5 and 6xHis Tags
US-10-865-978-37

Query Match 65.0%; Score 58.5; DB 16; Length 326;
Best Local Similarity 57.1%; Pred. No. 0.93;
Matches 16; Conservative 0; Mismatches 1; Indels 11; Gaps 2;

QY 1 KGEGLPIPNPLLRG 17
|||||
Db 293 KGEGLPIPNPLLRG 320

RESULT 5
US-10-467-243-26
; Sequence 26, Application US/10467243
; Publication No. US20040132971A1
; GENERAL INFORMATION:
; APPLICANT: Maxygen Holdings Ltd.
; APPLICANT: Haaning, Jesper Mortensen
; APPLICANT: Halkier, Torben
; TITLE OF INVENTION: RANK LIGAND-BINDING POLYPEPTIDES
; FILE REFERENCE: 0226wo310
; CURRENT APPLICATION NUMBER: US/10/467,243
; CURRENT FILING DATE: 2003-08-06
; PRIOR APPLICATION NUMBER: DK PA 2001 00214
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: US 60/267,843
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: DK PA 2001 00498
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/278,320
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: soluble hOPG polypeptide encoded by SEQ ID NO:25
US-10-467-243-26

Query Match 65.0%; Score 58.5; DB 16; Length 338;
Best Local Similarity 59.1%; Pred. No. 0.97;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 1 KGEGLPIPNPLLRG 17
|||||
Db 311 KGEGLPIPNPLLRG 332

RESULT 6
US-10-467-243-8
; Sequence 8, Application US/10467243
; Publication No. US20040132971A1
; GENERAL INFORMATION:
; APPLICANT: Maxygen Holdings Ltd.
; APPLICANT: Haaning, Jesper Mortensen
; APPLICANT: Halkier, Torben
; TITLE OF INVENTION: RANK LIGAND-BINDING POLYPEPTIDES
; FILE REFERENCE: 0226wo310
; CURRENT APPLICATION NUMBER: US/10/467,243
; CURRENT FILING DATE: 2003-08-06
; PRIOR APPLICATION NUMBER: DK PA 2001 00214
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: US 60/267,843
; PRIOR FILING DATE: 2001-02-09

```
; PRIOR APPLICATION NUMBER: DK PA 2001 00498
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/278,320
; PRIOR FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: polypeptide encoded by SEQ ID NO:7
US-10-467-243-8

Query Match      65.0%; Score 58.5; DB 16; Length 349;
Best Local Similarity 59.1%; Pred. No. 1;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

Qy      1 KGELEGLPIPNPLL-----RTG 17
Db      322 RGPFGKPIPNPLGLDSTRTG 343

RESULT 7
US-10-270-223-4
; Sequence 4, Application US/10270223
; Publication No. US20030143634A1
; GENERAL INFORMATION:
; APPLICANT: BioImage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS
; TITLE OF INVENTION: INTERACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION RE
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Aequoria Victoria and Human
US-10-270-223-4

Query Match      65.0%; Score 58.5; DB 14; Length 368;
Best Local Similarity 59.1%; Pred. No. 1.1;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

Qy      1 KGELEGLPIPNPLL-----RTG 17
Db      341 RGPFGKPIPNPLGLDSTRTG 362

RESULT 8
US-10-239-652A-6
; Sequence 6, Application US/10239652A
; Publication No. US20040053234A1
; GENERAL INFORMATION:
; APPLICANT: Michael David Winther; Heidi Lynn Smith; Andre Ponton;
; APPLICANT: Roberto Justo De Antueno; Stephen John Allen
; TITLE OF INVENTION: Polynucleotides that Control Delta-6-Desaturase Genes
; TITLE OF INVENTION: and Methods for Identifying Compounds for Modulating
; TITLE OF INVENTION: Delta-6-Desaturase
; FILE REFERENCE: 42320-0010
; CURRENT APPLICATION NUMBER: US/10/239,652A
; CURRENT FILING DATE: 2003-07-28
; PRIOR APPLICATION NUMBER: PCT/CA01/00398
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: CA2,301,158
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-415-232-4

Query Match      65.0%; Score 58.5; DB 15; Length 473;
Best Local Similarity 59.1%; Pred. No. 1.4;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;
```

```
; TYPE: PRT
; ORGANISM: rat
US-10-239-652A-6

Query Match      65.0%; Score 58.5; DB 15; Length 473;
Best Local Similarity 59.1%; Pred. No. 1.4;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

Qy      1 KGELEGLPIPNPLL-----RTG 17
Db      446 RGPFGKPIPNPLGLDSTRTG 467

RESULT 9
US-10-239-652A-8
; Sequence 8, Application US/10239652A
; Publication No. US20040053234A1
; GENERAL INFORMATION:
; APPLICANT: Michael David Winther; Heidi Lynn Smith; Andre Ponton;
; APPLICANT: Roberto Justo De Antueno; Stephen John Allen
; TITLE OF INVENTION: Polynucleotides that Control Delta-6-Desaturase Genes
; TITLE OF INVENTION: and Methods for Identifying Compounds for Modulating
; TITLE OF INVENTION: Delta-6-Desaturase
; FILE REFERENCE: 42320-0010
; CURRENT APPLICATION NUMBER: US/10/239,652A
; CURRENT FILING DATE: 2003-07-28
; PRIOR APPLICATION NUMBER: PCT/CA01/00398
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: CA2,301,158
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Human
US-10-239-652A-8

Query Match      65.0%; Score 58.5; DB 15; Length 473;
Best Local Similarity 59.1%; Pred. No. 1.4;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

Qy      1 KGELEGLPIPNPLL-----RTG 17
Db      446 RGPFGKPIPNPLGLDSTRTG 467

RESULT 10
US-10-415-232-4
; Sequence 4, Application US/10415232
; Publication No. US20040096435A1
; GENERAL INFORMATION:
; APPLICANT: Winther et al.
; TITLE OF INVENTION: Methods for Screening Compounds that Modulate Lipid
; TITLE OF INVENTION: Metabolism
; FILE REFERENCE: 04511.0005.NPUS01
; CURRENT APPLICATION NUMBER: US/10/415,232
; CURRENT FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: PCT/CA01/01520
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 60/243,009
; PRIOR FILING DATE: 2000-10-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-415-232-4

Query Match      65.0%; Score 58.5; DB 15; Length 473;
Best Local Similarity 59.1%; Pred. No. 1.4;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;
```


RESULT 15
US-10-653-496A-25
; Sequence 25, Application US/10653496A
; Publication No. US20040118672A1
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Li, Jinhe
; APPLICANT: Pauley, Adele M.
; APPLICANT: Pharmacia & Upjohn Company
; TITLE OF INVENTION: Human Sel-10 Polypeptides and Polynucleotides that
; FILE REFERENCE: 6142
; CURRENT APPLICATION NUMBER: US/10/653,496A
; PRIOR FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/09/328,877D
; PRIOR FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: V5HIS tagged
; OTHER INFORMATION: homo sapien
US-10-653-496A-25

Query Match 65.0%; Score 58.5; DB 16; Length 669;
Best Local Similarity 59.1%; Pred. No. 2.1;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

Qy 1 KGELEGPIPNPLL-----RTG 17
Db 642 RGPFGKPIPNPLLDSTRTG 663

Search completed: May 27, 2005, 14:15:04
Job time : 5.92964 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:28:05 ; Search time 6.45203 Seconds
(without alignments)
1019.048 Million cell updates/sec

Title: US-10-622-893A-4
Perfect score: 90
Sequence: 1 KGELEGLPIPNPLRTG 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59.5	66.1	32	8 ADN61411	Adn61411 Paenibaci
2	59	65.6	146	5 AAU76975	Aau76975 Human tra
3	58.5	65.0	42	8 ADN40846	Adn40846 Synthetic
4	58.5	65.0	250	4 AAB20357	Aab20357 Anti-chel
5	58.5	65.0	254	8 ADQ98103	Adq98103 Chimeric
6	58.5	65.0	254	8 ADQ98105	Adq98105 Chimeric
7	58.5	65.0	254	8 ADQ98104	Adq98104 Chimeric
8	58.5	65.0	254	8 ADQ98107	Adq98107 Chimeric
9	58.5	65.0	254	8 ADQ98106	Adq98106 Chimeric
10	58.5	65.0	254	8 ADQ98108	Adq98108 Chimeric
11	58.5	65.0	338	5 ABG71834	Abg71834 Soluble h
12	58.5	65.0	349	5 ABG71826	Abg71826 Protein e
13	58.5	65.0	368	6 ABR40351	AbR40351 Human ami
14	58.5	65.0	374	3 AAB10639	Aab10639 Human VEG
15	58.5	65.0	473	4 AAE11082	Aae11082 C-termina
16	58.5	65.0	473	4 AAE11084	Aae11084 C-termina
17	58.5	65.0	473	5 AAE14740	Aae14740 C-termina
18	58.5	65.0	669	2 AAY22470	Aay22470 Human Cte
19	58.5	65.0	669	4 AAB59202	Aab59202 C-termina
20	56.5	62.8	253	8 ADJ62976	Adj62976 Novel flu
21	56.5	62.8	361	6 ABR61836	AbR61836 P. falcip
22	56.5	62.8	383	5 ABB07412	Abb07412 Amino aci
23	56.5	62.8	383	6 ABG72053	Abg72053 Human Hsk
24	55.5	61.7	570	8 ADR86696	Adr86696 Ephrin B4
25	55.5	61.7	570	8 ADR82643	Adr82643 Human B4E

26	55.5	61.7	1620	8 ADH71972	Adh71972 Human pro
27	55.5	61.7	1653	8 ADH71970	Adh71970 Human pro
28	51.5	57.2	28	8 ADQ48584	Adq48584 Viral vec
29	51.5	57.2	33	8 ADQ48586	Adq48586 Viral vec
30	51.5	57.2	33	8 ADQ48573	Adq48573 Viral vec
31	51.5	57.2	42	8 ADQ48581	Adq48581 Viral vec
32	51.5	57.2	42	8 ADQ48577	Adq48577 Viral vec
33	51.5	57.2	42	8 ADQ48579	Adq48579 Viral vec
34	51.5	57.2	135	8 ADJ63148	Adj63148 Novel flu
35	51.5	57.2	181	8 ADJ63042	Adj63042 Novel flu
36	51.5	57.2	196	8 ADJ63116	Adj63116 Novel flu
37	51.5	57.2	200	8 ADJ63004	Adj63004 Novel flu
38	51.5	57.2	212	8 ADJ63024	Adj63024 Novel flu
39	51.5	57.2	233	8 ADJ63128	Adj63128 Novel flu
40	51.5	57.2	236	8 ADJ63012	Adj63012 Novel flu
41	51.5	57.2	237	8 ADJ63082	Adj63082 Novel flu
42	51.5	57.2	237	8 ADJ63058	Adj63058 Novel flu
43	51.5	57.2	237	8 ADJ63144	Adj63144 Novel flu
44	51.5	57.2	237	8 ADJ63022	Adj63022 Novel flu
45	51.5	57.2	237	8 ADJ63080	Adj63080 Novel flu

ALIGNMENTS

RESULT 1
ADN61411
ID ADN61411 standard; peptide; 32 AA.
XX
AC ADN61411;
XX
DT 01-JUL-2004 (first entry)
XX
DE Paenibacillus sp. Cry1529-related V5 Ab epitope/His tag purity peptide.
XX
KW Cry; toxic; lepidopteran pest; toxin complex; insecticide; DAS1529; ORF7;
KW Cry1529; purification; V5 antibody epitope; His tag.
XX
OS Synthetic.
OS Unidentified.
XX
FH Key Location/Qualifiers
FT Region 10..26
FT /note= "V5 antibody epitope"
XX
XX WO2004002223-A2.
PD 08-JAN-2004.
XX
XX 27-JUN-2003; 2003WO-US020082.
PF
XX 28-JUN-2002; 2002US-0392633P.
PR
XX 21-JAN-2003; 2003US-0441647P.
XX
XX (DOWC) DOW AGROSCIENCES LLC.
PA
XX
XX Bintrim SB, Bevan SA, Zhu B, Merlo DJ;
XX WPI; 2004-082821/08.
DR
XX
XX Screening a culture of Paenibacillus isolate for Cry protein or toxin
XX complex protein, useful for controlling lepidopterans, comprises
XX obtaining DNA or protein from the culture and assaying the presence of
XX the gene or protein.
XX
XX Example 11; Page 70; 220pp; English.
PS
XX
XX The invention relates to a novel method for screening a culture of a
XX Paenibacillus isolate for a gene encoding a protein selected from a Cry
XX protein that is toxic to a lepidopteran pest and a toxin complex protein.
XX The method comprises obtaining DNA from the culture and assaying the DNA
XX for the presence of the gene or obtaining a protein produced by the
XX culture and assaying the presence of a protein that indicates the

CC polypeptide is an artificial tag protein sequence for an RNA polymerase
 CC of the invention.
 XX
 SQ Sequence 42 AA;

Query Match 65.0%; Score 58.5; DB 8; Length 42;
 Best Local Similarity 59.1%; Pred. No. 0.086; Length 42;
 Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

Qy 1 KGELEGLPIPNPLL-----RTG 17
 :||| ||||| |||
 Db 15 RGFEGKPIPNPLLGLDSTRTG 36

RESULT 4
 AAB20357
 ID AAB20357 standard; protein; 250 AA.

AC AAB20357;

XX 11-JUN-2001 (first entry)

DE Anti-chelate antibody CHA255 Fab heavy chain.

XX Antibody engineering; metal chelate; CHA255; indium; EDTA; tumour;
 KW cancer; therapy.
 KW
 XX

OS Mus musculus.

XX Location/Qualifiers

PH Key 1..32
 FT Region /label= FR1
 FT /note= "framework region 1"
 FT 33..37
 FT Region /label= CDR1
 FT /note= "complementarity determining region 1"
 FT 38..51
 FT Region /label= FR2
 FT /note= "framework region 2"
 FT 53..68
 FT Region /label= CDR2
 FT /note= "complementarity determining region 2"
 FT 69..100
 FT Region /label= FR3
 FT /note= "framework region 3"
 FT 101..105
 FT Region /label= CDR3
 FT /note= "complementarity determining region 3"
 FT 106..116
 FT Region /label= FR4
 FT /note= "framework region 1"
 FT 245..250
 FT Peptide /label= Epitope_tag

XX WO200122922-A2.

PN 05-APR-2001.

XX 27-SEP-2000; 2000WO-US026619.

XX 27-SEP-1999; 99US-0156194P.

XX 31-MAY-2000; 2000US-0208684P.

XX (REGC) UNIV CALIFORNIA.

XX Meares C, Chmura A;

XX WPI; 2001-244971/25.

XX N-PSDB; AAF30632.

XX Nucleic acid encoding a mutant antibody comprising a reactive site that
 PT specifically binds to a metal chelate useful as analytical agents and in
 PT clinical diagnosis, as well as in the treatment of disease, particularly

PT cancer.

PS Example 4; Fig 15; 100pp; English.

XX The invention provides a mutant antibody comprising a reactive site that
 CC is not present in the wild-type of the antibody. The antibody also has a
 CC complementarity determining region (CDR) that specifically binds to a
 CC metal chelate against which the wild-type antibody is raised. The
 CC reactive site of the mutant antibody is in a position proximate to or
 CC within the CDR, such that the chelate and the antibody are able to form a
 CC covalent bond. The present sequence is that of anti-indium-EDTA
 CC monoclonal antibody CHA255 Fab heavy chain, including a C-terminal
 CC epitope tag. As an example of the method of the invention, rational
 CC computer-aided design was used to develop an indium-EDTA chelate to
 CC covalently bind to CHA255 in vivo. The premise was to allow the chelate
 CC to bind non-covalently to CHA255 bound to a tumour and then to covalently
 CC attach the chelate to the antibody, thereby trapping it at the tumour
 CC site. This involved cloning the variable domains of CHA255 to construct a
 CC human/mouse chimeric Fab fragment that could be expressed in Escherichia
 CC coli, and the synthesis and screening of benzyl-EDTA chelates carrying
 CC weakly electrophilic groups capable of conjugation of the antibody in
 CC vivo. This Fab can be conjugated to a targeting moiety when desired. A
 CC reactive site was incorporated into the antibody by engineering a Cys
 CC residue at location Asn-96 or Ser-95 of the light chain, near the region
 CC of the antibody to which the chelate bound. Such engineering is typically
 CC accomplished by site-directed mutagenesis of a nucleic acid encoding the
 CC wild-type of the anti-chelate antibody
 XX

SQ Sequence 250 AA;

Query Match 65.0%; Score 58.5; DB 4; Length 250;

Best Local Similarity 59.1%; Pred. No. 0.057;

Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

Qy 1 KGELEGLPIPNPLL-----RTG 17

Db 223 RGFEGKPIPNPLLGLDSTRTG 244

RESULT 5

ADQ98103

ID ADQ98103 standard; protein; 254 AA.

XX ADQ98103;

XX 21-OCT-2004 (first entry)

DE Chimeric murine 2D12.5 VH antibody fused to human TTCH protein with tag.

XX murine; mouse; chimeric; human; TTCH; tetanus toxin;

KW magnetic resonance imaging; lanthanide luminescence; gamma-emissions;

KW single photon emission tomography; SPET; cancer; cytostatic;

KW immunosuppressive; multi-functional antibody; metal chelate;

KW antigen recognition domain; in vivo imaging;

KW cell-antibody-metal chelate complex; emission tomography; V5 epitope;

KW His tag.

OS Mus musculus.

OS Homo sapiens.

OS Chimeric.

OS Synthetic.

XX WO2004065569-A2.

XX 05-AUG-2004.

XX 23-JAN-2004; 2004WO-US001808.

XX 23-JAN-2003; 2003US-00350555.

XX 22-JUL-2003; 2003US-00625047.

XX 31-JUL-2003; 2003US-00631258.

XX (REGC) UNIV CALIFORNIA.

```

XX Meares C, Corneillie T;
XX WPI; 2004-580725/56.
XX N-PSDB; ADQ98096.
XX
XX Novel mutant antibody comprising reactive site not present in wild-type
XX of antibody and antigen recognition domain that recognizes macrocyclic
XX metal chelate having four nitrogen atoms, useful for treating cancer or
XX autoimmune diseases.
XX
XX Claim 41; SEQ ID NO 80; 208pp; English.
XX
XX This invention relates to multi-functional antibodies that recognise
XX chelating agents and metal chelates, particularly macrocyclic metal
XX chelates. Specifically, it refers to an antibody that comprises a metal
XX chelate bound to an antigen recognition domain, where the metal chelate
XX has a reactive functional group of complementary reactivity to the
XX reactive site of the antibody. This reactive site is the side chain of a
XX naturally occurring amino acid e.g. the -SH group side chain of a
XX cysteine residue (not present in the wild type) which can be used to form
XX a covalent bond between the reactive site of the antibody and the
XX reactive functional group of the metal chelate. The present invention
XX describes using these antibodies for in vivo imaging where the antibody
XX comprises a targeting moiety that binds specifically to a cell via a cell
XX surface receptor or antigen thus forming a cell-mutant antibody complex.
XX On addition of the metal chelate, a cell-antibody-metal chelate complex
XX is formed that can be detected using emission tomography, magnetic
XX resonance imaging, lanthanide luminescence, gamma-emissions or single
XX photon emission tomography (SPET). As such, this method is useful for
XX treating a subject with cancer and pharmaceutical compositions exhibit
XX cytostatic and immunosuppressive activities. This polypeptide sequence is
XX a chimeric murine antibody 2D12.5 variable heavy chain protein fused to
XX the human anti-tetanus toxin antibody CH1 heavy chain constant region
XX with a V5 epitope and (His)6 tag, given in an exemplification of the
XX invention.
XX
XX Sequence 254 AA;
XX
XX Query Match 65.0%; Score 58.5; DB 8; Length 254;
XX Best Local Similarity 59.1%; Pred. No. 0.58;
XX Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;
XX
XX QY 1 KGELEGLPIPNPLL-----RTG 17
XX Db 227 RGFPEKGKPIPNPLLDLSDTRTG 248
XX
XX RESULT 6
XX ID ADQ98105
XX AC ADQ98105 standard; protein; 254 AA.
XX AC ADQ98105;
XX
XX 21-OCT-2004 (first entry)
XX
XX Chimeric 2dVH-TTCH protein with tag and an N87D/G53C mutations SeqID 83.
XX
XX murine; mouse; chimeric; human; TTCH; tetanus toxin;
XX magnetic resonance imaging; lanthanide luminescence; gamma-emissions;
XX single photon emission tomography; SPET; cancer; cytostatic;
XX immunosuppressive; multi-functional antibody; metal chelate;
XX antigen recognition domain; in vivo imaging;
XX cell-antibody-metal chelate complex; emission tomography; V5 epitope;
XX His tag; mutant; mutein.
XX
XX Mus musculus.
XX OS Homo sapiens.
XX OS Chimeric.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 53
XX FT

```

```

FT Meares C, Corneillie T;
FT Misc-difference 87
FT WPI; 2004-580725/56.
FT N-PSDB; ADQ98096.
FT
FT Novel mutant antibody comprising reactive site not present in wild-type
FT of antibody and antigen recognition domain that recognizes macrocyclic
FT metal chelate having four nitrogen atoms, useful for treating cancer or
FT autoimmune diseases.
FT
FT Claim 41; SEQ ID NO 82; 208pp; English.
FT
FT This invention relates to multi-functional antibodies that recognise
FT chelating agents and metal chelates, particularly macrocyclic metal
FT chelates. Specifically, it refers to an antibody that comprises a metal
FT chelate bound to an antigen recognition domain, where the metal chelate
FT has a reactive functional group of complementary reactivity to the
FT reactive site of the antibody. This reactive site is the side chain of a
FT naturally occurring amino acid e.g. the -SH group side chain of a
FT cysteine residue (not present in the wild type) which can be used to form
FT a covalent bond between the reactive site of the antibody and the
FT reactive functional group of the metal chelate. The present invention
FT describes using these antibodies for in vivo imaging where the antibody
FT comprises a targeting moiety that binds specifically to a cell via a cell
FT surface receptor or antigen thus forming a cell-mutant antibody complex.
FT On addition of the metal chelate, a cell-antibody-metal chelate complex
FT is formed that can be detected using emission tomography, magnetic
FT resonance imaging, lanthanide luminescence, gamma-emissions or single
FT photon emission tomography (SPET). As such, this method is useful for
FT treating a subject with cancer and pharmaceutical compositions exhibit
FT cytostatic and immunosuppressive activities. This polypeptide sequence is
FT a chimeric murine antibody 2D12.5 variable heavy chain protein fused to
FT the human anti-tetanus toxin antibody CH1 heavy chain constant region
FT with a V5 epitope and (His)6 tag, given in an exemplification of the
FT invention.
FT
FT Sequence 254 AA;
FT
FT Query Match 65.0%; Score 58.5; DB 8; Length 254;
FT Best Local Similarity 59.1%; Pred. No. 0.58;
FT Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;
FT
FT QY 1 KGELEGLPIPNPLL-----RTG 17
FT Db 227 RGFPEKGKPIPNPLLDLSDTRTG 248
FT
FT RESULT 7
FT ID ADQ98104
FT AC ADQ98104 standard; protein; 254 AA.
FT AC ADQ98104;
FT
FT 21-OCT-2004 (first entry)
FT
FT Chimeric 2dVH-TTCH protein with tag and an N87D mutation SeqID 82.
FT

```

XX murine; mouse; chimeric; human; TTCH; tetanus toxin;
KW magnetic resonance imaging; lanthanide luminescence; gamma-emissions;
KW single photon emission tomography; SPET; cancer; cytostatic;
KW immunosuppressive; multi-functional antibody; metal chelate;
KW antigen recognition domain; in vivo imaging;
KW cell-antibody-metal chelate complex; emission tomography; V5 epitope;
KW His tag; mutant; mutein.
XX
OS Mus musculus.
OS Homo sapiens.
OS Chimeric.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 87 /note= "Wild type Asn substituted for Asp"
XX
XX PN WO2004065569-A2.
XX PD 05-AUG-2004.
XX PF 23-JAN-2004; 2004WO-US001808.
XX PR 23-JAN-2003; 2003US-00350555.
XX PR 22-JUL-2003; 2003US-00625047.
XX PR 31-JUL-2003; 2003US-00631258.
XX
XX PA (REGC) UNIV CALIFORNIA.
XX PI Meares C, Corneillie T;
XX
XX DR WPI; 2004-580725/56.
XX DR N-PSDB; ADQ98097.
XX
XX PT Novel mutant antibody comprising reactive site not present in wild-type
PT of antibody and antigen recognition domain that recognizes macrocyclic
PT metal chelate having four nitrogen atoms, useful for treating cancer or
PT autoimmune diseases.
XX
XX PS Claim 41; SEQ ID NO 81; 208pp; English.
XX
XX CC This invention relates to multi-functional antibodies that recognise
CC chelating agents and metal chelates, particularly macrocyclic metal
CC chelates. Specifically, it refers to an antibody that comprises a metal
CC chelate bound to an antigen recognition domain, where the metal chelate
CC has a reactive functional group of complementary reactivity to the
CC reactive site of the antibody. This reactive site is the side chain of a
CC naturally occurring amino acid e.g. the -SH group side chain of a
CC cysteine residue (not present in the wild type) which can be used to form
CC a covalent bond between the reactive site of the antibody and the
CC reactive functional group of the metal chelate. The present invention
CC describes using these antibodies for in vivo imaging where the antibody
CC comprises a targeting moiety that binds specifically to a cell via a cell
CC surface receptor or antigen thus forming a cell-mutant antibody complex.
CC On addition of the metal chelate, a cell-antibody-metal chelate complex
CC is formed that can be detected using emission tomography, magnetic
CC resonance imaging, lanthanide luminescence, gamma-emissions or single
CC photon emission tomography (SPET). As such, this method is useful for
CC treating a subject with cancer and pharmaceutical compositions exhibit
CC cytostatic and immunosuppressive activities. This polypeptide is the
CC mutant protein sequence comprising a chimeric murine antibody 2D12.5
CC variable heavy chain protein fused to the human anti-tetanus toxin
CC antibody CH1 heavy chain constant region (2dVH-TTCH) with an N87D
CC mutation and a V5 epitope with (His)6 tag, given in an exemplification of
CC the invention.
XX
SQ Sequence 254 AA;

Query Match 65.0%; Score 58.5; DB 8; Length 254;
Best Local Similarity 59.1%; Pred. No. 0.58;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 1 KGELEGLPIPNPLL-----RTG 17
DB 227 RGPFGKPIPNPLLGLDSTRTG 248
RESULT 8
ADQ98107
ID ADQ98107 standard; protein; 254 AA.
XX AC ADQ98107;
XX AC AC
XX XX 21-OCT-2004 (first entry)
XX DE Chimeric 2dVH-TTCH protein with tag and an N87D/G55C mutations SeqID 85.
XX KW murine; mouse; chimeric; human; TTCH; tetanus toxin;
KW magnetic resonance imaging; lanthanide luminescence; gamma-emissions;
KW single photon emission tomography; SPET; cancer; cytostatic;
KW immunosuppressive; multi-functional antibody; metal chelate;
KW antigen recognition domain; in vivo imaging;
KW cell-antibody-metal chelate complex; emission tomography; V5 epitope;
KW His tag; mutant; mutein.
XX
OS Mus musculus.
OS Homo sapiens.
OS Chimeric.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 55 /note= "Wild type Gly substituted for Cys"
FT Misc-difference 87 /note= "Wild type Asn substituted for Asp"
XX
XX PN WO2004065569-A2.
XX PD 05-AUG-2004.
XX PF 23-JAN-2004; 2004WO-US001808.
XX PR 23-JAN-2003; 2003US-00350555.
XX PR 22-JUL-2003; 2003US-00625047.
XX PR 31-JUL-2003; 2003US-00631258.
XX
XX PA (REGC) UNIV CALIFORNIA.
XX PI Meares C, Corneillie T;
XX
XX DR WPI; 2004-580725/56.
XX DR N-PSDB; ADQ98100.
XX
XX PT Novel mutant antibody comprising reactive site not present in wild-type
PT of antibody and antigen recognition domain that recognizes macrocyclic
PT metal chelate having four nitrogen atoms, useful for treating cancer or
PT autoimmune diseases.
XX
XX PS Claim 41; SEQ ID NO 84; 208pp; English.
XX
XX CC This invention relates to multi-functional antibodies that recognise
CC chelating agents and metal chelates, particularly macrocyclic metal
CC chelates. Specifically, it refers to an antibody that comprises a metal
CC chelate bound to an antigen recognition domain, where the metal chelate
CC has a reactive functional group of complementary reactivity to the
CC reactive site of the antibody. This reactive site is the side chain of a
CC naturally occurring amino acid e.g. the -SH group side chain of a
CC cysteine residue (not present in the wild type) which can be used to form
CC a covalent bond between the reactive site of the antibody and the
CC reactive functional group of the metal chelate. The present invention
CC describes using these antibodies for in vivo imaging where the antibody
CC comprises a targeting moiety that binds specifically to a cell via a cell
CC surface receptor or antigen thus forming a cell-mutant antibody complex.
CC On addition of the metal chelate, a cell-antibody-metal chelate complex
CC is formed that can be detected using emission tomography, magnetic

CC resonance imaging, lanthanide luminescence, gamma-emissions or single
 CC photon emission tomography (SPET). As such, this method is useful for
 CC treating a subject with cancer and pharmaceutical compositions exhibit
 CC cytostatic and immunosuppressive activities. This polypeptide is the
 CC mutant protein sequence comprising a chimeric murine antibody 2D12.5
 CC variable heavy chain protein fused to the human anti-tetanus toxin
 CC antibody CH1 heavy chain constant region (2dVH-TTCH) with N87D and G55C
 CC mutations and a V5 epitope with (His)6 tag, given in an exemplification
 CC of the invention.
 XX
 SQ Sequence 254 AA;

Query Match 65.0%; Score 58.5; DB 8; Length 254;
 Best Local Similarity 59.1%; Pred. NO. 0.58;
 Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 1 KGELEGLPIPNPLL-----RTG 17
 Db 227 RGPFEKPIPNPLGLDSTRTG 248

RESULT 9
 ADQ98106
 ID ADQ98106 standard; protein; 254 AA.
 XX
 AC ADQ98106;

XX
 DT 21-OCT-2004 (first entry)

XX Chimeric 2dVH-TTCH protein with tag and an N87D/G54C mutations SeqID 84.
 DE
 KW murine; mouse; chimeric; human; TTCH; tetanus toxin;
 KW magnetic resonance imaging; lanthanide luminescence; gamma-emissions;
 KW single photon emission tomography; SPET; cancer; cytostatic;
 KW immunosuppressive; multi-functional antibody; metal chelate;
 KW antigen recognition domain; in vivo imaging;
 KW cell-antibody-metal chelate complex; emission tomography; V5 epitope;
 KW His tag; mutant; mutein.

XX Mus musculus.
 OS Homo sapiens.
 OS Chimeric.
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 54 /note= "Wild type Gly substituted for Cys"

FT Misc-difference 87 /note= "Wild type Asn substituted for Asp"

FT

XX WO2004065569-A2.

XX

XX

XX

XX

XX 23-JAN-2004; 2004WO-US001808.

XX

XX 23-JAN-2003; 2003US-00350555.

XX 22-JUL-2003; 2003US-00625047.

XX 31-JUL-2003; 2003US-00631258.

XX (REGC) UNIV CALIFORNIA.

XX

XX Meares C, Corneillie T;

XX

XX WPI; 2004-580725/56.

XX N-PSDB; ADQ98099.

XX Novel mutant antibody comprising reactive site not present in wild-type
 PT of antibody and antigen recognition domain that recognizes macrocyclic
 PT metal chelate having four nitrogen atoms, useful for treating cancer or
 PT autoimmune diseases.

XX Claim 41; SEQ ID NO 83; 208pp; English.

PS

XX This invention relates to multi-functional antibodies that recognise
 CC chelating agents and metal chelates, particularly macrocyclic metal
 CC chelates. Specifically, it refers to an antibody that comprises a metal
 CC chelate bound to an antigen recognition domain, where the metal chelate
 CC has a reactive functional group of complementary reactivity to the
 CC reactive site of the antibody. This reactive site is the side chain of a
 CC naturally occurring amino acid e.g. the -SH group side chain of a
 CC cysteine residue (not present in the wild type) which can be used to form
 CC a covalent bond between the reactive site of the antibody and the
 CC reactive functional group of the metal chelate. The present invention
 CC describes using these antibodies for in vivo imaging where the antibody
 CC comprises a targeting moiety that binds specifically to a cell via a cell
 CC surface receptor or antigen thus forming a cell-mutant antibody complex.
 CC On addition of the metal chelate, a cell-antibody-metal chelate complex
 CC is formed that can be detected using emission tomography, magnetic
 CC resonance imaging, lanthanide luminescence, gamma-emissions or single
 CC photon emission tomography (SPET). As such, this method is useful for
 CC treating a subject with cancer and pharmaceutical compositions exhibit
 CC cytostatic and immunosuppressive activities. This polypeptide is the
 CC mutant protein sequence comprising a chimeric murine antibody 2D12.5
 CC variable heavy chain protein fused to the human anti-tetanus toxin
 CC antibody CH1 heavy chain constant region (2dVH-TTCH) with N87D and G54C
 CC mutations and a V5 epitope with (His)6 tag, given in an exemplification
 CC of the invention.
 XX
 SQ Sequence 254 AA;

Query Match 65.0%; Score 58.5; DB 8; Length 254;
 Best Local Similarity 59.1%; Pred. NO. 0.58;
 Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 1 KGELEGLPIPNPLL-----RTG 17
 Db 227 RGPFEKPIPNPLGLDSTRTG 248

RESULT 10

ADQ98108

ID ADQ98108 standard; protein; 254 AA.

XX

AC ADQ98108;

XX

DT 21-OCT-2004 (first entry)

XX

DE Chimeric murine 2D12.5 VH antibody fused to human TTCH protein with tag.

XX

KW murine; mouse; chimeric; human; TTCH; tetanus toxin;

KW magnetic resonance imaging; lanthanide luminescence; gamma-emissions;

KW single photon emission tomography; SPET; cancer; cytostatic;

KW immunosuppressive; multi-functional antibody; metal chelate;

KW antigen recognition domain; in vivo imaging;

KW cell-antibody-metal chelate complex; emission tomography; V5 epitope;

KW His tag.

XX

OS Mus musculus.

OS Homo sapiens.

OS Chimeric.

OS Synthetic.

XX WO2004065569-A2.

XX

XX 05-AUG-2004.

XX

XX 23-JAN-2004; 2004WO-US001808.

XX

XX 23-JAN-2003; 2003US-00350555.

XX 22-JUL-2003; 2003US-00625047.

XX 31-JUL-2003; 2003US-00631258.

XX (REGC) UNIV CALIFORNIA.

XX Meares C, Corneillie T;

PI

PA	(JANC) JANSSEN PHARM NV.
XX	
PI	Gordon RD, Sprengel JJ, Yon JR, Dijkmans JJH, Gosiewska A;
PI	Dhanaraj SN, Xu J;
XX	
XX	
DR	WPI; 2000-442669/38.
DR	N-PSDB; AAA71983.
XX	
PT	New vascular endothelial growth factor protein, useful for treating or
PT	preventing diseases associated with inappropriate angiogenesis activity
PT	such as cancer, rheumatoid arthritis, psoriasis and wounds.
XX	
PS	Disclosure; Fig 19; 127pp; English.
XX	
CC	This invention describes a novel vascular endothelial growth factor-X
CC	(VEGF-X) protein (Ia) and its encoding polynucleotide (IIa) which has
CC	vulnerary, cytostatic, antirheumatic, antiarthritic, antipsoriatic and
CC	antidiabetic activity and acts as an angiogenesis and vasculization
CC	regulator. An antisense molecule of the invention is useful for treating
CC	or preventing cancer, rheumatoid arthritis, psoriasis and diabetic
CC	retinopathy by inhibiting angiogenic activity or inappropriate
CC	vascularization including formation and proliferation of new blood
CC	vessels, growth and development of tissues, tissue regeneration and organ
CC	and tissue repair in a subject. The products of the invention are useful
CC	for preparing medicaments for treating wounds such as dermal ulcers,
CC	pressure sores, venous sores, diabetic ulcers and burns and to promote
CC	skin graft growth, tissue repair, proliferation of new blood vessels,
CC	tissue regeneration and organ repair by promoting angiogenic activity or
CC	vascularization. This sequence represents a human VEGF-X protein which
CC	can be expressed in mammalian systems and which is described in the
CC	method of the invention
XX	
SQ	Sequence 374 AA;
	Query Match 65.0%; Score 58.5; DB 3; Length 374;
	Best Local Similarity 59.1%; Pred.No. 0.87;
	Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;
QY	1 KGELEGLIPNPILL-----RTG 17
	:
Db	347 RGPFEKPIPPLGLDSTRTG 368
	:
RESULT 15	
AAE11082	
ID	AAE11082 standard; protein; 473 AA.
AC	AAE11082;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	C-terminal tagged rat delta-6-desaturase (rD6D-1).
XX	
KW	Delta-6-desaturase gene; D6D; lipid metabolism disorder; atopic eczema;
KW	mastalgia; rheumatoid arthritis; Sjogren's syndrome; viral infection;
KW	gastrointestinal disorder; post viral fatigue; pre-menstrual syndrome;
KW	endometriosis; cystic fibrosis; alcoholism; Alzheimer's disease;
KW	cardiovascular disease; Crohn's disease; congenital liver disease;
KW	schizophrenia; diabetic neuropathy; nephropathy; retinopathy; cancer;
KW	arterial hypertension; atherosclerosis; chronic inflammatory disorder;
KW	autoimmune disorder; hypercholesterolaemia; atopic disorder; rD6D-1;
KW	gene therapy; rat.
OS	Rattus sp.
XX	
FH	Key Location/Qualifiers
FT	Region 451..464
FT	/note= "V5 tag"
FT	Region 468..473
FT	/note= "6xHis tag"
XX	
XX	WO200170993-A2.
PX	
XX	

PD	27-SEP-2001.
XX	
PF	26-MAR-2001; 2001WO-CA000398.
XX	
PR	24-MAR-2000; 2000CA-02301158.
XX	
FA	(SCOT-) SCOTIA HOLDINGS PLC.
XX	
PI	Winther MD, Smith HL, Allen SJ, Ponton A, De Antueno RJ;
XX	
DR	WPI; 2001-611507/70.
XX	
FT	Nucleic acid encoding delta-6-desaturase gene useful for treating atopic
PT	eczema, mastalgia, rheumatoid arthritis, Sjogren's syndrome,
PT	gastrointestinal disorders, viral infections and post viral fatigue.
XX	
PS	Example 3; Fig 5; 164pp; English.
XX	
CC	The invention relates to polynucleotides that control delta-6 desaturase
CC	genes (D6D) and methods useful for identifying compounds which inhibit or
CC	promote the activity of mammalian D6D. Compounds which modulate D6D gene
CC	segments are useful for treating lipid metabolism disorders e.g. atopic
CC	eczema, mastalgia, rheumatoid arthritis, Sjogren's syndrome, e.g. atopic
CC	gastrointestinal disorders, viral infections and post viral fatigue, pre-
CC	menstrual syndrome, endometriosis, cystic fibrosis, alcoholism,
CC	Alzheimer's syndrome, cardiovascular disease, Crohn's disease, cancer,
CC	congenital liver disease, schizophrenia, diabetes and diabetic
CC	complications including diabetic neuropathy, nephropathy and retinopathy.
CC	Compounds of the invention are also useful for inhibiting progressive and
CC	acute disorders such as arterial hypertension, atherosclerosis and other
CC	inflammatory and autoimmune disorders, hypercholesterolaemia and other
CC	atopic disorders. D6D genes are useful in gene therapy. The present
CC	sequence is C-terminal tagged rat delta-6-desaturase (rD6D-1) enzyme
XX	
SQ	Sequence 473 AA;
	Query Match 65.0%; Score 58.5; DB 4; Length 473;
	Best Local Similarity 59.1%; Pred.No. 1.1;
	Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1
QY	1 KGELEGLIPNPILL-----RTG 17
	:
Db	446 RGPFEKPIPPLGLDSTRTG 467
	:
	Search completed: May 27, 2005, 13:43:56
	Job time : 8.45203 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:35:35 ; Search time 1.3049 Seconds
(without alignments)
972.511 Million cell updates/sec

Title: US-10-622-893A-4
Perfect score: 90
Sequence: 1 KGELEGPIPNPLRTG 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgm2_6/ptodata/1/iaa/5A-COMB.pep:*
2: /cgm2_6/ptodata/1/iaa/5B-COMB.pep:*
3: /cgm2_6/ptodata/1/iaa/6A-COMB.pep:*
4: /cgm2_6/ptodata/1/iaa/6B-COMB.pep:*
5: /cgm2_6/ptodata/1/iaa/PCTUS-COMB.pep:*
6: /cgm2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	65.6	146	4	US-09-923-304-9
2	58.5	65.0	374	4	US-09-468-647A-118
3	58.5	65.0	669	4	US-09-213-888-25
4	58.5	65.0	669	4	US-09-328-877D-25
5	56.5	62.8	383	4	US-09-883-096-5
6	51.5	57.2	635	4	US-09-545-814-32
7	45	50.0	758	1	US-07-756-250-16
8	45	50.0	766	4	US-09-949-016-10058
9	44	48.9	410	4	US-09-489-039A-10689
10	44	48.9	1013	2	US-08-866-650-3
11	44	48.9	1013	2	US-09-021-287-3
12	44	48.9	1013	3	US-09-240-473-3
13	43	47.8	14	4	US-09-352-171-14
14	43	47.8	622	4	US-09-949-016-9104
15	42	46.7	271	4	US-09-270-767-60778
16	42	46.7	313	4	US-09-902-540-13536
17	42	46.7	348	4	US-09-489-039A-8827
18	42	46.7	393	4	US-09-252-991A-20362
19	42	46.7	473	3	US-08-857-076-99
20	42	46.7	593	3	US-08-991-408-4
21	42	46.7	591	3	US-09-432-473-4
22	42	46.7	778	4	US-10-148-806-36
23	42	46.7	1013	2	US-08-866-650-5
24	42	46.7	1013	2	US-09-021-287-5
25	42	46.7	1013	3	US-08-991-408-2
26	42	46.7	1013	3	US-09-240-473-5
27	42	46.7	1013	3	US-09-432-473-2

RESULT 1
US-09-923-304-9
; Sequence 9, Application US/09923304
; Patent No. 6797471
; GENERAL INFORMATION:
; APPLICANT: KATZ, RUTH
; APPLICANT: JIANG, FENG
; TITLE OF INVENTION: DETECTION AND DIAGNOSIS OF SMOKING RELATED CANCERS
; FILE REFERENCE: UTSC:6580S
; CURRENT APPLICATION NUMBER: US/09/923.304
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-923-304-9

Query Match 65.6%; Score 59; DB 4; Length 146;
Best Local Similarity 59.3%; Pred. No. 0.06;
Matches 16; Conservative 0; Mismatches 1; Indels 10; Gaps 2;
Qy 1 KGEI-----EGLPIPNPL-----RTG 17
Db 114 KGEI--GHPFEGKPIPNPLGLDSTRTG 140
RESULT 2
US-09-468-647A-118
; Sequence 118, Application US/09468647A
; Patent No. 6783953
; GENERAL INFORMATION:
; APPLICANT: Gordon, Robert D
; APPLICANT: Sprengel, Jorg J
; APPLICANT: Yen, Jeffery R
; APPLICANT: Dijkmans, Josiena J.H.
; APPLICANT: Gosciewska, Anna
; APPLICANT: Dhanaraj, Sri devi N
; APPLICANT: Xu, Jean
; TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-X
; FILE REFERENCE: B0192.70011US00
; CURRENT APPLICATION NUMBER: US/09/468.647A
; CURRENT FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: GB 9828377.3
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: US 60/124,967

ALIGNMENTS

28 42 46.7 1013 4 US-09-285-385C-20 Sequence 20, Appl
29 42 46.7 1013 4 US-09-902-540-15897 Sequence 15897, A
30 41 45.6 232 4 US-09-134-000C-4038 Sequence 4038, Ap
31 41 45.6 429 4 US-09-902-540-12898 Sequence 12898, A
32 41 45.6 430 4 US-10-132-920B-2 Sequence 2, Appl
33 41 45.6 430 4 US-09-700-696C-2 Sequence 2, Appl
34 41 45.6 435 4 US-09-270-767-45936 Sequence 45936, A
35 41 45.6 509 4 US-09-794-422-34 Sequence 34, Appl
36 41 45.6 514 4 US-09-292-858B-14 Sequence 14, Appl
37 41 45.6 525 4 US-10-132-920B-27 Sequence 27, Appl
38 41 45.6 525 4 US-09-794-422-6 Sequence 6, Appl
39 41 45.6 540 4 US-09-794-422-46 Sequence 46, Appl
40 41 45.6 556 4 US-09-794-422-8 Sequence 8, Appl
41 45.6 559 4 US-10-116-370-2 Sequence 2, Appl
42 41 45.6 567 4 US-09-949-016-10952 Sequence 10952, A
43 41 45.6 706 4 US-09-248-796A-18837 Sequence 18837, A
44 41 45.6 772 4 US-10-148-806-37 Sequence 37, Appl
45 41 45.6 1509 4 US-09-677-046A-2 Sequence 2, Appl

```
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: US 60/164,131
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 118
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-468-647A-118

Query Match      65.0%; Score 58.5; DB 4; Length 374;
Best Local Similarity 59.1%; Pred. No. 0.2;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 1 KGEGLPIPNPLL-----RTG 17
Db 347 RGFEGKPIPNPLGLDSTRTG 368

RESULT 3
US-09-213-888-25
; Sequence 25, Application US/09213888A
; Patent No. 6638731
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Li, Jinhe
; APPLICANT: Pauley, Adele M.
; APPLICANT: Pharmacia & Upjohn Company
; TITLE OF INVENTION: Human Sel-10 Polypeptides and Polynucleotides that
; FILE OF INVENTION: Encode Them
; FILE REFERENCE: 6142
; CURRENT APPLICATION NUMBER: US/09/213,888A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: VSHIS tagged
; OTHER INFORMATION: homo sapien
US-09-213-888-25

Query Match      65.0%; Score 58.5; DB 4; Length 669;
Best Local Similarity 59.1%; Pred. No. 0.39;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 1 KGEGLPIPNPLL-----RTG 17
Db 642 RGFEGKPIPNPLGLDSTRTG 663

RESULT 4
US-09-328-877D-25
; Sequence 25, Application US/09328877D
; Patent No. 6730778
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Li, Jinhe
; APPLICANT: Pauley, Adele M.
; APPLICANT: Pharmacia & Upjohn Company
; TITLE OF INVENTION: Human Sel-10 Polypeptides and Polynucleotides that
; FILE OF INVENTION: Encode Them
; FILE REFERENCE: 6142
; CURRENT APPLICATION NUMBER: US/09/328,877D
; CURRENT FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Tagged Ctenocephalides felis

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: VSHIS tagged
; OTHER INFORMATION: homo sapien
US-09-328-877D-25

Query Match      65.0%; Score 58.5; DB 4; Length 669;
Best Local Similarity 59.1%; Pred. No. 0.39;
Matches 13; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 1 KGEGLPIPNPLL-----RTG 17
Db 642 RGFEGKPIPNPLGLDSTRTG 663

RESULT 5
US-09-883-096-5
; Sequence 5, Application US/09883096
; Patent No. 6680369
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Craven, Andrew
; APPLICANT: Yu, Ming
; APPLICANT: Sakowicz, Roman
; APPLICANT: Patel, Umesh A.
; APPLICANT: Davies, Katherine A.
; TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR USE
; FILE REFERENCE: 020552-001410US
; CURRENT APPLICATION NUMBER: US/09/883,096
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 09/594,655
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Hskip3a
; OTHER INFORMATION: fragment
; OTHER INFORMATION: Amino acid sequence of Hskip3a fragment used in
; OTHER INFORMATION: the ATPase assay (Figure 4).
US-09-883-096-5

Query Match      62.8%; Score 56.5; DB 4; Length 383;
Best Local Similarity 65.0%; Pred. No. 0.44;
Matches 13; Conservative 1; Mismatches 1; Indels 5; Gaps 1;

QY 3 ELEGLPINPLL-----RTG 17
Db 358 KLEGKPIPNPLGLDSTRTG 377

RESULT 6
US-09-545-814-32
; Sequence 32, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
; APPLICANT: Becher, Anna M.
; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE OF INVENTION: USSES THEREOF
; FILE REFERENCE: FC-5-C1
; CURRENT APPLICATION NUMBER: US/09/545,814
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,833
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 635
; TYPE: PRT
; ORGANISM: Tagged Ctenocephalides felis
```

US-09-545-814-32

Query Match 57.2%; Score 51.5; DB 4; Length 635;
Best Local Similarity 66.7%; Pred. No. 4.8;
Matches 12; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

Qy 5 EGLPIPNPLG-----RTG 17
Db 612 EGKPIPNPLGLDSTRIG 629

RESULT 7

US-07-756-250-16
; Sequence 16, Application US/07756250
; Patent No. 5268275
; GENERAL INFORMATION:
; APPLICANT: Wu, Sheue-Mei
; APPLICANT: Stafford, Darrel W.
; TITLE OF INVENTION: Vitamin K-Dependent Carboxylase
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and
; ADDRESSEE: Gibson
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5268275th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07756,250
; FILING DATE: 19910909
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/697,427
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-34
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102

INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 758 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-756-250-16

Query Match 50.0%; Score 45; DB 1; Length 758;
Best Local Similarity 50.0%; Pred. No. 63;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KGELEGLPIPNPLRT 16
Db 643 EGEVKGPEPTPLVQT 658

RESULT 8

US-09-949-016-10058
; Sequence 10058, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10058
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10058

Query Match 50.0%; Score 45; DB 4; Length 766;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KGELEGLPIPNPLRT 16
Db 651 EGEVKGPEPTPLVQT 666

RESULT 9

US-09-489-039A-10689
; Sequence 10689, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10689
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-10689

Query Match 48.9%; Score 44; DB 4; Length 410;
Best Local Similarity 58.3%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 6 GLPIPNPLRTG 17
Db 323 GLPLPQPTVETG 334

RESULT 10

US-08-866-650-3
; Sequence 3, Application US/08866650
; Patent No. 5939321
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S

; APPLICANT: Takahara, Kazuhiko
; APPLICANT: Hoffman, Guy G
; TITLE OF INVENTION: Mammalian Tolloid-Like Protein
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703

COMPUTER READABLE FORM:

```
/ ; MEDIUM TYPE: Floppy disk
/ ; COMPUTER: IBM PC compatible
/ ; OPERATING SYSTEM: PC-DOS/MS-DOS
/ ; SOFTWARE: PatentIn Release #1.0, Version #1.30
/ ; CURRENT APPLICATION DATA:
/ ; APPLICATION NUMBER: US/08/866,650
/ ; FILING DATE:
/ ; CLASSIFICATION: 514
/ ; ATTORNEY/AGENT INFORMATION:
/ ; NAME: Berson, Bennett J
/ ; REGISTRATION NUMBER: 37094
/ ; REFERENCE/DOCKET NUMBER: 960296.93839
/ ; TELEPHONE: 608-251-5000
/ ; TELEFAX: 608-251-9166
/ ; INFORMATION FOR SEQ ID NO: 3:
/ ; LENGTH: 1013 amino acids
/ ; TYPE: amino acid
/ ; TOPOLOGY: linear
/ ; MOLECULE TYPE: protein
/ ; US-08-866-650-3
/ ;
/ ; Query Match 48.9%; Score 44; DB 2; Length 1013;
/ ; Best Local Similarity 56.2%; Pred. No. 1.2e+02;
/ ; Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
/ ;
/ ; QY 2 GELEGLPIPNPLRTG 17
/ ; Db 846 GRLCGSKIPLMATG 861
/ ;
/ ; RESULT 11
/ ; US-09-021-287-3
/ ; Sequence 3, Application US/09021287
/ ; Patent No. 5981717
/ ; GENERAL INFORMATION:
/ ; APPLICANT: Greenspan, Daniel S
/ ; APPLICANT: Takahara, Kazuhiko
/ ; APPLICANT: Hoffman, Guy G
/ ; TITLE OF INVENTION: Mammalian Tolloid-Like Protein
/ ; NUMBER OF SEQUENCES: 13
/ ; CORRESPONDENCE ADDRESS:
/ ; ADDRESSEE: Quarles & Brady
/ ; STREET: 1 South Pinckney Street
/ ; CITY: Madison
/ ; STATE: WI
/ ; COUNTRY: US
/ ; ZIP: 53703
/ ; COMPUTER READABLE FORM:
/ ; MEDIUM TYPE: Floppy disk
/ ; COMPUTER: IBM PC compatible
/ ; OPERATING SYSTEM: PC-DOS/MS-DOS
/ ; SOFTWARE: PatentIn Release #1.0, Version #1.30
/ ; CURRENT APPLICATION DATA:
/ ; APPLICATION NUMBER: US/09/021,287
/ ; FILING DATE:
/ ; CLASSIFICATION:
/ ; ATTORNEY/AGENT INFORMATION:
/ ; NAME: Berson, Bennett J
/ ; REGISTRATION NUMBER: 37094
/ ; REFERENCE/DOCKET NUMBER: 960296.93839
/ ; TELEPHONE: 608-251-5000
/ ; TELEFAX: 608-251-9166
/ ; INFORMATION FOR SEQ ID NO: 3:
/ ; LENGTH: 1013 amino acids
/ ; TYPE: amino acid
/ ; TOPOLOGY: linear
/ ; MOLECULE TYPE: protein
/ ; US-09-021-287-3
/ ;
/ ; Query Match 48.9%; Score 44; DB 2; Length 1013;
/ ; Best Local Similarity 56.2%; Pred. No. 1.2e+02;
/ ; Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
/ ;
/ ; QY 2 GELEGLPIPNPLRTG 17
/ ; Db 846 GRLCGSKIPLMATG 861
/ ;
/ ; RESULT 12
/ ; US-09-240-473-3
/ ; Sequence 3, Application US/09240473
/ ; Patent No. 6297011
/ ; GENERAL INFORMATION:
/ ; APPLICANT: Greenspan, Daniel S
/ ; APPLICANT: Takahara, Kazuhiko
/ ; APPLICANT: Hoffman, Guy G
/ ; TITLE OF INVENTION: Mammalian Tolloid-Like Protein
/ ; NUMBER OF SEQUENCES: 13
/ ; CORRESPONDENCE ADDRESS:
/ ; ADDRESSEE: Quarles & Brady
/ ; STREET: 1 South Pinckney Street
/ ; CITY: Madison
/ ; STATE: WI
/ ; COUNTRY: US
/ ; ZIP: 53703
/ ; COMPUTER READABLE FORM:
/ ; MEDIUM TYPE: Floppy disk
/ ; COMPUTER: IBM PC compatible
/ ; OPERATING SYSTEM: PC-DOS/MS-DOS
/ ; SOFTWARE: PatentIn Release #1.0, Version #1.30
/ ; CURRENT APPLICATION DATA:
/ ; APPLICATION NUMBER: US/09/240,473
/ ; FILING DATE:
/ ; CLASSIFICATION:
/ ; ATTORNEY/AGENT INFORMATION:
/ ; NAME: Berson, Bennett J
/ ; REGISTRATION NUMBER: 37094
/ ; REFERENCE/DOCKET NUMBER: 960296.93839
/ ; TELEPHONE: 608-251-5000
/ ; TELEFAX: 608-251-9166
/ ; INFORMATION FOR SEQ ID NO: 3:
/ ; SEQUENCE CHARACTERISTICS:
/ ; LENGTH: 1013 amino acids
/ ; TYPE: amino acid
/ ; TOPOLOGY: linear
/ ; MOLECULE TYPE: protein
/ ; US-09-240-473-3
/ ;
/ ; Query Match 48.9%; Score 44; DB 3; Length 1013;
/ ; Best Local Similarity 56.2%; Pred. No. 1.2e+02;
/ ; Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
/ ;
/ ; QY 2 GELEGLPIPNPLRTG 17
/ ; Db 846 GRLCGSKIPLMATG 861
/ ;
/ ; RESULT 13
/ ; US-09-352-171-14
/ ; Sequence 14, Application US/09352171
/ ; Patent No. 6759206
/ ; GENERAL INFORMATION:
/ ; APPLICANT: Rubin, Richard A
/ ; APPLICANT: Conway, Bruce
/ ; APPLICANT: Giuliani, Kenneth A
/ ; APPLICANT: Gough, Albert H
/ ; APPLICANT: Dunlay, R. Terry
/ ; TITLE OF INVENTION: A System for Cell Based Screening
```

```
/ ; MOLECULE TYPE: protein
/ ; US-09-021-287-3
/ ;
/ ; Query Match 48.9%; Score 44; DB 2; Length 1013;
/ ; Best Local Similarity 56.2%; Pred. No. 1.2e+02;
/ ; Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
/ ;
/ ; QY 2 GELEGLPIPNPLRTG 17
/ ; Db 846 GRLCGSKIPLMATG 861
/ ;
/ ; RESULT 12
/ ; US-09-240-473-3
/ ; Sequence 3, Application US/09240473
/ ; Patent No. 6297011
/ ; GENERAL INFORMATION:
/ ; APPLICANT: Greenspan, Daniel S
/ ; APPLICANT: Takahara, Kazuhiko
/ ; APPLICANT: Hoffman, Guy G
/ ; TITLE OF INVENTION: Mammalian Tolloid-Like Protein
/ ; NUMBER OF SEQUENCES: 13
/ ; CORRESPONDENCE ADDRESS:
/ ; ADDRESSEE: Quarles & Brady
/ ; STREET: 1 South Pinckney Street
/ ; CITY: Madison
/ ; STATE: WI
/ ; COUNTRY: US
/ ; ZIP: 53703
/ ; COMPUTER READABLE FORM:
/ ; MEDIUM TYPE: Floppy disk
/ ; COMPUTER: IBM PC compatible
/ ; OPERATING SYSTEM: PC-DOS/MS-DOS
/ ; SOFTWARE: PatentIn Release #1.0, Version #1.30
/ ; CURRENT APPLICATION DATA:
/ ; APPLICATION NUMBER: US/09/240,473
/ ; FILING DATE:
/ ; CLASSIFICATION:
/ ; ATTORNEY/AGENT INFORMATION:
/ ; NAME: Berson, Bennett J
/ ; REGISTRATION NUMBER: 37094
/ ; REFERENCE/DOCKET NUMBER: 960296.93839
/ ; TELEPHONE: 608-251-5000
/ ; TELEFAX: 608-251-9166
/ ; INFORMATION FOR SEQ ID NO: 3:
/ ; SEQUENCE CHARACTERISTICS:
/ ; LENGTH: 1013 amino acids
/ ; TYPE: amino acid
/ ; TOPOLOGY: linear
/ ; MOLECULE TYPE: protein
/ ; US-09-240-473-3
/ ;
/ ; Query Match 48.9%; Score 44; DB 3; Length 1013;
/ ; Best Local Similarity 56.2%; Pred. No. 1.2e+02;
/ ; Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
/ ;
/ ; QY 2 GELEGLPIPNPLRTG 17
/ ; Db 846 GRLCGSKIPLMATG 861
/ ;
/ ; RESULT 13
/ ; US-09-352-171-14
/ ; Sequence 14, Application US/09352171
/ ; Patent No. 6759206
/ ; GENERAL INFORMATION:
/ ; APPLICANT: Rubin, Richard A
/ ; APPLICANT: Conway, Bruce
/ ; APPLICANT: Giuliani, Kenneth A
/ ; APPLICANT: Gough, Albert H
/ ; APPLICANT: Dunlay, R. Terry
/ ; TITLE OF INVENTION: A System for Cell Based Screening
```

; FILE REFERENCE: 97,022-D1
; CURRENT APPLICATION NUMBER: US/09/352,171
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: epitope tag
US-09-352-171-14

Query Match 47.8%; Score 43; DB 4; Length 14;
Best Local Similarity 88.9%; Pred. No. 1.6;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 GLPIPNPLL 14
| | | | | | | |
Db 1 GKPIPNPLL 9

RESULT 14
US-09-949-016-9104
; Sequence 9104, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9104
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9104

Query Match 47.8%; Score 43; DB 4; Length 622;
Best Local Similarity 58.3%; Pred. No. 1.1e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 6 GLPIPNLLRTG 17
| | | | | | | |
Db 14 GYPLGFLAQTG 25

RESULT 15
US-09-270-767-60778
; Sequence 60778, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60778
; LENGTH: 271
; TYPE: PRT
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-60778

Query Match 46.7%; Score 42; DB 4; Length 271;
Best Local Similarity 57.1%; Pred. No. 61;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 GELEGLPIPNPLLR 15
| | | | | | | |
Db 229 GEIDGLTIPKNLLQ 242

Search completed: May 27, 2005, 13:50:00
Job time : 2.3049 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:28:55 ; Search time 3.73348 Seconds
(without alignments)
2331.700 Million cell updates/sec

Title: US-10-622-893A-4
Perfect score: 90
Sequence: 1 KGEGLPIPNPLRTG 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	55.6	131	2 QBJW23	QBJW23 human adeno
2	50	55.6	131	2 QBB8S5	QBB8S5 human adeno
3	50	55.6	529	1 FTSK_HAEIN	P45264 haemophilus
4	48	53.3	131	2 Q67729	Q67729 human adeno
5	48	53.3	154	2 QBP406	QBP406 xanthomonas
6	47	52.2	147	2 Q94HT5	Q94HT5 oryza sativ
7	47	52.2	147	2 Q7XG14	Q7XG14 oryza sativ
8	47	52.2	358	1 C8BA_RAT	P05554 rattus norv
9	47	52.2	522	2 Q7QG23	Q7QG23 anopheles g
10	47	52.2	529	1 Y243_ARCFU	O29996 archaeoglob
11	47	52.2	579	2 Q94HM1	Q94HM1 oryza sativ
12	47	52.2	579	2 Q7XG15	Q7XG15 oryza sativ
13	47	52.2	2039	2 Q7S3G9	Q7S3G9 neurospora
14	46	51.1	202	2 Q6MND1	Q6MND1 bdellovibri
15	46	51.1	255	2 Q8UG13	Q8UG13 agrobacteri
16	46	51.1	310	2 Q8A5C1	Q8A5C1 bacteroides
17	46	51.1	1494	2 Q6CG13	Q6CG13 yarrowia li
18	45	50.0	166	2 Q6P9X3	Q6P9X3 rattus norv
19	45	50.0	261	2 Q8REU8	Q8REU8 fusobacteri
20	45	50.0	263	2 Q7P3V5	Q7P3V5 fusobacteri
21	45	50.0	299	2 Q6GL93	Q6GL93 xenopus tro
22	45	50.0	372	2 Q8PZL8	Q8PZL8 brucella su
23	45	50.0	378	2 Q8YI97	Q8YI97 brucella me
24	45	50.0	560	2 Q9YIV7	Q9YIV7 pseudophila
25	45	50.0	645	2 Q88RK0	Q88RK0 pseudomonas
26	45	50.0	757	1 VKGC_MOUSE	Q9GYC7 mus musculu
27	45	50.0	758	1 VKGC_BOVIN	Q07175 bos taurus
28	45	50.0	758	1 VKGC_HUMAN	P38435 homo sapien
29	45	50.0	758	1 VKGC_RAT	Q88496 rattus norv
30	45	50.0	758	2 Q9GL59	Q9GL59 ovis aries
31	45	50.0	758	2 Q9MYY3	Q9MYY3 delphinapte

32	45	50.0	930	1 FTSK_PASMU	Q9CPI3 pasteurella
33	45	50.0	1138	2 Q6BZA3	Q6BZA3 debaryomyce
34	45	50.0	1371	2 P73337	P73337 synchocyst
35	44.5	49.4	356	2 Q7UYG2	Q7UYG2 rhodospirell
36	44	48.9	103	1 HEX9_ADECT	P14268 canine aden
37	44	48.9	103	2 Q7M6A0	Q7M6A0 canine aden
38	44	48.9	145	2 Q67106	Q67106 aquifex seo
39	44	48.9	257	1 TRMD_BORBR	Q7U384 bordetella
40	44	48.9	257	1 TRMD_BORPA	Q7U374 bordetella
41	44	48.9	257	1 TRMD_BORPE	Q7U374 bordetella
42	44	48.9	266	2 Q54360	Q54360 streptomyce
43	44	48.9	275	2 Q67SK7	Q67SK7 symbiobacte
44	44	48.9	339	2 Q8TQC9	Q8TQC9 methanosarc
45	44	48.9	482	2 Q8PA76	Q8PA76 xanthomonas

Q8JW23	PRELIMINARY;	PRT;	131 AA.
AC	Q8JW23;		
DT	01-OCT-2002 (Tremblrel. 22, Created)		
DT	01-OCT-2002 (Tremblrel. 22, Last sequence update)		
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)		
DB	E3 14.9kDa.		
OS	Human adenovirus type 34.		
OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.		
OX	NCBI_TaxID=10548;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Compton;		
RX	PubMed=14693847;		
RA	Adhikary A.K., Inada T., Numaga J., Suzuki E., Ushijima H., Banik U.,		
RA	Mukoyama A., Matsuno S., Okabe N.;		
RT	"Characterization of hexon and fiber genes of a novel strain of		
RT	adenovirus involved in epidemic keratoconjunctivitis.";		
RL	J. Clin. Pathol. 57:95-97(2004).		
DR	EMBL; AB079724; BAC07480.1; -.		
DR	InterPro; IPR009266; Adeno_E3.		
DR	Pfam; PF06040; Adeno_E3; 1.		
SQ	SEQUENCE 131 AA; 14966 MW; 6EAE3911AADS1FBD CRC64;		
Query Match	55.6%; Score 50; DB 2; Length 131;		
Best Local Similarity	56.2%; Pred. No. 6.6;		
Matches	9; Conservative	2; Mismatches	5; Indels
			0; Gaps
			0;

Qy	2 GELEGLPIPNPLRTG 17		
Db	77 GELEGLPIPNPLRTG 92		

Q8B8S5	PRELIMINARY;	PRT;	131 AA.
AC	Q8B8S5;		
DT	01-MAR-2003 (Tremblrel. 23, Created)		
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)		
DT	05-JUN-2004 (Tremblrel. 27, Last annotation update)		
DE	E3 14.6 kDa protein (16.1K protein) (E3 CRI-alpha)		
GN	Name=E3A;		
OS	Human adenovirus type 11.		
OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.		
OX	NCBI_TaxID=10541;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Adilp Slobitski;		
RX	MEDLINE=22614046; PubMed=12726735; DOI=10.1016/S0042-6822(02)00085-5;		
RA	Stone D., Furthmann A., Sandig V., Lieber A.;		
RT	"The complete nucleotide sequence, genome organization, and origin of		
RT	human adenovirus type 11.";		
RL	Virology 309:152-165(2003).		

ALIGNMENTS

```

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Slobitski;
RX MEDLINE=22749818; PubMed=12867636; DOI=10.1099/vir.0.19178-0;
RA Mei Y.F., Skog J., Lindman K., Wadell G.;
RT "Comparative analysis of the genome organization of human adenovirus
RT 11, a member of the human adenovirus species B, and the commonly used
RT human adenovirus 5 vector, a member of species C.";
RL J. Gen. Virol. 84:2061-2071(2003).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Slobitski;
RA Mei Y.F., Skog J., Lindman K., Wadell G.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22935380; PubMed=14573794; DOI=10.1099/vir.0.19497-0;
RA Davison A.J., Benko M., Harrach B.;
RT "Genetic content and evolution of adenoviruses.";
RL J. Gen. Virol. 84:2895-2908(2003).
DR EMBL; AV163756; AAN62494.1; -.
DR EMBL; AF532578; AAP49187.1; -.
DR EMBL; BK001453; DAA01668.1; -.
DR InterPro; IPR009266; Adeno_E3.
DR Pfam; PF06040; Adeno_E3; 1.
SQ SEQUENCE 131 AA; 14966 MW; 6EAE3911AAD51FBD CRC64;

Query Match 55.6%; Score 50; DB 2; Length 131;
Best Local Similarity 56.2%; Pred. No. 6.6;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 GELEGLPIPNPLRLTG 17
| | | | | | | | | |
DB 77 GELEGLPIPNPLRLTG 92

RESULT 3
FTSK_HAEIN
ID_FTSK_HAEIN STANDARD; PRT; 529 AA.
AC P45264;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE DNA translocase ftsk.
GN Name=ftsK; OrderedLocusNames=HI1592;
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G.G., FitzHugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: Probable DNA motor protein. May track DNA in a ATP-
CC dependent manner by generating positive supercoils in front of it
CC and negative supercoils behind it (By similarity).
CC -1- MISCELLANEOUS: Although strongly related to ftsK, it lacks the
CC typical transmembrane domains located at the N-terminal part of
CC ftsK.
CC -1- SIMILARITY: Contains 1 FtsK domain.
-----
```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32833; AAC23240.1; -.
DR PIR; G64131; G64131.
DR TIGR; H11592; -.
DR HAMAP; MF_01809; atypical; 1.
DR InterPro; IPR002543; FtsK_SpoIIIE.
DR InterPro; IPR009058; Wing_hlx_DNA_bnd.
DR Pfam; PF01580; FtsK_SpoIIIE; 1.
DR PROSITE; PS0901; FtsK; 1.
KW ATP-binding; Cell division; Chromosome partition; Complete proteome;
KW DNA-binding.
FT NP_BIND 174 387 FtsK.
FT DOMAIN 191 198 ATP (Potential).
SQ SEQUENCE 529 AA; 58767 MW; 0C0F882EADFE726D CRC64;

Query Match 55.6%; Score 50; DB 1; Length 529;
Best Local Similarity 53.3%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 ELEGPLIPNPLRLTG 17
| | | | | | | | | |
DB 286 EAMGMFVPNPIWRLG 300

RESULT 4
Q67729
ID Q67729 PRELIMINARY; PRT; 131 AA.
AC Q67729; Q7T929;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE 15.0 kDa protein (E3 15.0K protein) (16.1K protein).
GN Name=E3A;
OS Human adenovirus type 35.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=10522;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Holden;
RX MEDLINE=89012230; PubMed=3172347;
RA Flomenberg P.R., Chen M., Horwitz M.S.;
RT "Sequence and genetic organization of adenovirus type 35 early region
RT 3.";
RL J. Virol. 62:4431-4437(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Holden;
RA Basler C.F., Flomenberg P.F., Chen M., Horwitz M.S.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC VR-718;
RX MEDLINE=22890078; PubMed=14528318;
RA Gao W., Robbins P.D., Gambotto A.;
RT "Human adenovirus type 35: nucleotide sequence and vector
RT development.";
RL Gene Ther. 10:1941-1949(2003).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=35p;
RX MEDLINE=22741997; PubMed=12857895;
RX DOI=10.1128/JVI.77.15.8263-8271.2003;
RA Vogels R., Zuijdgeest D., Van Rijnsvoever R., Hartkoorn E., Damen I.,
RA De Bechune M.P., Kostense S., Penders G., Helmus N., Koudstaal W.,
RA Cecchini M., Wetterwald A., Sprangers M., Lemckert A., Ophorst O.,
RA Koel B., Van Meerendonk M., Quax P., Panitti L., Grimbergen J.,
-----
```

```

RA  Bout A., Goudsmit J., Havenga M.;
RT  "Replication-deficient human adenovirus type 35 vectors for gene
RT  transfer and vaccination: efficient human cell infection and bypass of
RL  J. Virol. 77:8263-8271(2003).
DR  EMBL; AY128640; AAN17492.1; -.
DR  EMBL; AY128640; AAN17492.1; -.
DR  EMBL; AY271307; AAP22358.1; -.
DR  PIR; JC4767; JC4767.
DR  InterPro; IPR009266; Adeno_E3.
DR  Pfam; PF06040; Adeno_E3; 1.
SQ  SEQUENCE 131 AA; 14974 MW; E4869B881FC9F743 CRC64;

Query Match      53.3%; Score 48; DB 2; Length 131;
Best Local Similarity 56.2%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY  2 GELEGLPIPNPLRTG 17
    |||||
DB  77 GELHGLTFNPWVEAG 92

RESULT 5
Q8P4Q6 PRELIMINARY; PRT; 154 AA.
AC Q8P4Q6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein XCC3651.
GN OrderedLocusNames=XCC3651.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Cipina L.P.,
RA Ciccarelli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE012486; AAM42921.1; -.
DR HSP; P14930; l1lD.
DR InterPro; IPR002579; MsrB.
DR InterPro; IPR011057; Msr4_like.
DR Pfam; PF01641; Selr; 1.
DR ProDom; PD004057; DUF25; 1.
DR TIGRFAMs; TIGR00357; DUF25; 1.
KW Complete proteome.
SQ SEQUENCE 154 AA; 16794 MW; 44390679394D7D1F CRC64;

Query Match      53.3%; Score 48; DB 2; Length 154;
Best Local Similarity 61.5%; Pred. No. 16;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY  5 EGLPIPNPLRTG 17
    :|||:
DB  77 GELHGLTFNPWVEAG 92

```

```

Db 136 DQQLPPLQAG 148

RESULT 6
Q94HT5 PRELIMINARY; PRT; 147 AA.
AC Q94HT5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative retroelement.
GN Name=OSJNBb0036B06.22;
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=Nipponbare;
RA Du H., Minx P., Abbott A., Doebber A., de la Bastide M., Spiegel L.,
RA Nacimento L., Preston R., Kirchoff K., King L., Vil M.D., Baker J.,
RA Zutavern T., Santos L., Bell M., Miller B., Kuit K., Rodriguez S.,
RA Cunniss D.M., Baliya V., Shah R., Bahret A., O'Shaughnessy A.,
RA Palmer L., Yang C., Dedhia N., McCombie W.R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC051624; AAK92564.1; -.
DR Gramene; Q94HT5; -.
SQ SEQUENCE 147 AA; 16177 MW; 7CF958DBE5551325 CRC64;

Query Match      52.2%; Score 47; DB 2; Length 147;
Best Local Similarity 60.0%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  1 KGELEGLPIPNPLR 15
    |||||
DB  35 KGLGLPVRSPTLR 49

RESULT 7
Q7XG14 PRELIMINARY; PRT; 147 AA.
AC Q7XG14;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative retroelement.
GN ORFNames=OSJNBb0036B06.22;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN 1;
RP SEQUENCE FROM N.A.
RC The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RT chromosome 10."
RL Science 300:1566-1569(2003).
RN 2;
RP SEQUENCE FROM N.A.
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017073; AAP52859.1; -.
DR Gramene; Q7XG14; -.
SQ SEQUENCE 147 AA; 16177 MW; 7CF958DBE5551325 CRC64;

Query Match      52.2%; Score 47; DB 2; Length 147;
Best Local Similarity 60.0%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  1 KGELEGLPIPNPLR 15
    |||||
DB  35 KGLGLPVRSPTLR 49

```

Db 35 KGGLRGLPVRSP TLR 49

RESULT 8

CEBP_KAI	CEBP_RAT	STANDARD;	PRT;	358 AA.
ID	AC	P05554;		
DT	DT	01-NOV-1988 (Rel. 09, Created)		
DT	DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	DT	25-JAN-2005 (Rel. 46, Last annotation update)		
DE	DE	CCAAT/enhancer binding protein alpha (C/EBP		
GN	GN	Name=Cebpa;		
OS	OS	Rattus norvegicus (Rat).		
OC	OC	Eukaryota; Metazoa; Chordata; Craniata; Vert		
OC	OC	Mammalia; Eutheria; Rodentia; Sciurognathi;		
OX	OX	NCBI_TaxID=10116;		
RN	RN	[1]		
RP	RP	SEQUENCE FROM N.A., AND SEQUENCE OF 253-269.		
RC	RC	STRAIN=Sprague-Dawley;		
RX	RX	MEDLINE=89092001; PubMed=2850264;		
RA	RA	Landschulz W.H., Johnson P.F., Adashi E.Y.,		
RA	RA	McKnight S.L.;		
RT	RT	"Isolation of a recombinant copy of the gene		
RL	RL	Genes Dev. 2:786-800(1989).		

SEQUENCE FROM N. A. J. AND REVISIONS.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=95011606; PubMed=7926792;
RA Lincoln A.J., Williams S.C., Johnson P.F.;
RT "A revised sequence of the rat c/ebp gene.";
RL Genes Dev. 8:1131-1132(1994).
CC -!- FUNCTION: C/EBP is a DNA-binding protein
CC that forms a heterotrimeric complex with
CC the enhanced core homology common to many
CC different motifs: the CCAAT homology com-
CC mon to many enhancers; the C/EBP homology
CC common to many enhancers; and the C/EBP
CC homology common to many enhancers.
CC -!- SUBUNIT: Binds DNA as a dimer and can fo-
CC rm a heterotrimeric complex with C/EBP beta and gamma.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the bZIP family.
CC -----
CC This SWISS-PROT entry is copyright. It is pro-
CC tected by the Swiss Institute of Bioinformatics
CC and the European Bioinformatics Institute. Ther-
CC efore, this entry may not be reproduced or
CC used by non-profit institutions as long as
CC the source is acknowledged. This statement is not removed,
CC modified and this statement is not removed, as
CC the EMBL database requires a license agreement (See h-
CC ttp://www.ebi.ac.uk/EMBL/seqdb/doc/seqdb-sib.ch) or
CC send an email to: license@embl-sib.ch).

CC
DB EMBL: X12752: CAA31242.1: -

DR PIR; S06890; A54265.
DR TWO: Y 2000 2/C-270 340

DR TRANSFAC; T00108; -.

DR InterPro; IPR004827; TF bZIP.

DR SMARI; SM00338; BRGZ; I.
DR PROSTTE: PS50217: BZTP: 1

DR PROSITE; PS00036; BZIP_BASIC;

KW Nuclear protein; Transcription

FT	DOMAIN	262	272
FT	DOMAIN	262	272

FT	DNA BIND	286	306
FT	DOMAIN	317	345

SQ SEQUENCE 358 AA; 37371 MW;

Query Match 52.2%; S

Matches 10: Conservative 2:

Qv 2 GELEGI.PT PNBRL.PTG 17

[illegible]

RESULT 9

```

ID Q7Q3Z3 PRELIMINARY; PERT; 522 AA.
AC Q7Q3Z3;
DT DT 01-WAR-2004 (TREMBLrel. 26, Created)
DT DT 01-WAR-2004 (TREMBLrel. 26, Last sequence update)
DT DT 01-WAR-2004 (TREMBLrel. 26, Last annotation update)
DE AGCFL0524.
GN Name=agCG54096; ORFNames=ENSGG0000000010013;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
OX [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAB01008823; EAA05550.1; -.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR002401; EP450I.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00463; EP450I.
DR PRINTS; PR00385; P450.
DR DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
DR Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 522 AA; 59676 MW; 28E604C091E88DE9 CRC64;

```

Query Match 52.2%; Score 47; DB 2; Length 522;

BEST LOCAL SIMILARITY 33.3%; PIED: NO. 87;
Matches 8: Conservative 3: Mismatches 4: Indels 0: Gaps 0:

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

22

RESULT 10					
Y243 ARCFU					
ID Y243 ARCFU	STANDARD;	PRT;	529 AA.		
AC O29996;					
DT 15-JUL-1998 (Rel. 36, Created)					
DT 15-JUL-1998 (Rel. 36, Last sequence update)					
DT 25-OCT-2004 (Rel. 45, Last annotation update)					
DE Hypothetical protein AF0243.					
GN OrderedLocusNames=AF0243;					
OS Archaeoglobus fulgidus.					
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;					
OC Archaeoglobaceae; Archaeoglobus.					

```

RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001089; AAB90990.1; -.
DR PIR; C69280; C69280.
DR TIGR; AF0243; -.
DR InterPro; IPR006638; Elp3/Miab/NiFB.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF04055; Radical_SAM; 1.
DR SMART; SM00729; Elp3; 1.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 529 AA; 60503 MW; 57E9FBC2D0D284BBF CRC64;
SQ
Query Match 52.2%; Score 47; DB 1; Length 529;
Best Local Similarity 72.7%; Pred. No. 88;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 2 GELEGLPIPNP 12
Db 222 GEFEGVPRNP 232
||| ||| ||| |||
RESULT 11
Q94HML PRELIMINARY; PRT; 579 AA.
AC Q94HML;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative retroelement.
GN Name=OSUNBa0089D15.27;
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nipponbare;
RA Spiegel L., de la Bastide M., Nascimento L., Kirchhoff K., King L.,
RA Preston R., Vil M.D., Baker J., Bell M., Zutavern T., Santos L.,
RA Miller B., Kuit K., Rodriguez S., Cunnius D.M., Balija V., Shah R.,
RA Bahret A., Bai H., O'Shaughnessy A., Dedhia N., McCombie W.R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC078944; AAK92616.1; -.
DR Gramene; Q94HML; -.
DR InterPro; IPR009057; Homeodomain_like.
DR InterPro; IPR005162; Retrotrans_gag.
DR Pfam; PF03732; Retrotrans_gag; 1.
SQ SEQUENCE 579 AA; 63293 MW; 1832D98496BE25BF CRC64;
Query Match 52.2%; Score 47; DB 2; Length 579;
Best Local Similarity 60.0%; Pred. No. 97;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Qy 1 KGELEGLPIPNPLR 15
Db 35 KGLGLPVRSPTLR 49
||| ||| ||| |||
RESULT 12
Q7XG15 PRELIMINARY; PRT; 579 AA.
AC Q7XG15;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

```

```

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative retroelement.
GN ORFNames=OSUNBa0089D15.27;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC The Rice Chromosome 10 Sequencing Consortium;
RA "In-depth view of structure, activity, and evolution of rice
RT chromosome 10.";
RL Science 300:1566-1569(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017061; AAP52277.1; -.
DR Gramene; Q7XG15; -.
DR InterPro; IPR009057; Homeodomain_like.
DR InterPro; IPR005162; Retrotrans_gag.
DR Pfam; PF03732; Retrotrans_gag; 1.
SQ SEQUENCE 579 AA; 63293 MW; 1832D98496BE25BF CRC64;
Query Match 52.2%; Score 47; DB 2; Length 579;
Best Local Similarity 60.0%; Pred. No. 97;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Qy 1 KGELEGLPIPNPLR 15
Db 35 KGLGLPVRSPTLR 49
||| ||| ||| |||
RESULT 13
Q7S3G9 PRELIMINARY; PRT; 2039 AA.
AC Q7S3G9;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU06884.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rahman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Iankiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamvysellis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofora S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; ABE01000409; EAA29995.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR000379; Ser_estrs.
KW Hypothetical protein.

```

SQ SEQUENCE 2039 AA; 220637 MW; 62E6EAA1E02A38A3 CRC64;
Query Match 52.2%; Score 47; DB 2; Length 2039;
Best Local Similarity 64.3%; Pred. No. 3.7e+02;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 4 LEGLPIPNPLRTG 17
Db 1178 LDGLEIPTPALETG 1191
RESULT 14
Q6MND1 PRELIMINARY; PRT; 202 AA.
AC Q6MND1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=Bdi1327;
OS Bdellovibrio bacteriovorus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;
OC Bdellovibrionaceae; Bdellovibrio.
OX NCBI_TaxID=959;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;
RX PubMed=14752164; DOI=10.1126/science.1093027;
RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,
RA Keller H., Lambert C., Evans K.U., Goessmann A., Meyer F.,
RA Sockett R.E., Schuster S.C.;
RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a
genomic perspective."
RL Science 303:689-692(2004)
DR EMBL; BX842649; CAE79221.1; -.
KW Complete proteome.
SQ SEQUENCE 202 AA; 22573 MW; 35DBF670C9A4972E CRC64;

Query Match 51.1%; Score 46; DB 2; Length 202;
Best Local Similarity 58.3%; Pred. No. 45;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 GELEGLPIPNPL 13
Db 147 GFEGGLPVGNPI 158

RESULT 15
Q8UGI3 PRELIMINARY; PRT; 255 AA.
AC Q8UGI3; Q7D026;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein Atui054 (AGR_C1946p).
GN OrderedLocusNames=AGR_C1946, Atui054;
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Dupont;
RX MEDLINE=21608550; PubMed=11743193; DOI=10.1126/science.1066804;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,

RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Cereon;
RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Leppas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AE009069; AAL42067.1; -.
DR EMBL; AE008035; AAK86862.1; -.
DR PIR; AE2706; AE2706.
DR PIR; E97488; E97488.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 255 AA; 27291 MW; 5B329E22B9C43683 CRC64;
Query Match 51.1%; Score 46; DB 2; Length 255;
Best Local Similarity 42.9%; Pred. No. 58;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 KGELEGLPIPNPL 14
Db 68 EGDSDGIGLPDPLV 81

Search completed: May 27, 2005, 13:47:30
JOB time : 6.73348 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:34:21 ; Search time 1.3049 Seconds
(without alignments)
1253.491 Million cell updates/sec

Title: US-10-622-893A-4
Perfect score: 90
Sequence: 1 KGEGLPIPNPLRTG 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	55.6	529	2 G64131	sporulation protei
2	48	53.3	131	2 JC4767	E3 gene encoding h
3	47	52.2	358	2 A54265	CCAAT/enhancer-bin
4	47	52.2	529	2 C69280	conserved hypotet
5	46	51.1	255	2 AB2706	conserved hypotet
6	46	51.1	255	2 E97488	hypothetical prote
7	45	50.0	378	2 AF3320	flavohemoprotein [
8	45	50.0	560	2 S51600	phosphorylase kina
9	45	50.0	757	2 A39283	gamma-glutamyl car
10	45	50.0	758	2 A47439	gamma-glutamyl car
11	45	50.0	1371	2 S77521	sensory transducti
12	44	48.9	103	1 SXADC2	hexon-associated p
13	44	48.9	145	2 G70384	hypothetical prote
14	44	48.9	266	2 S44954	lmbG protein - Str
15	43.5	48.3	360	2 H87699	conserved hypotet
16	43.5	48.3	768	2 T17422	disease resistance
17	43	47.8	175	2 H72538	hypothetical prote
18	43	47.8	222	1 MNZSP	nonstructural prot
19	43	47.8	358	2 H75264	hypothetical prote
20	43	47.8	392	1 RRNZSP	polymerase-associa
21	43	47.8	820	2 T45685	hypothetical prote
22	43	47.8	855	2 B75191	probable ATP-depen
23	42.5	47.2	437	2 C39135	hypothetical prote
24	42	46.7	253	2 C81670	undecaprenyl pyrop
25	42	46.7	253	2 E71512	hypothetical prote
26	42	46.7	319	2 AG0700	probable pathogeni
27	42	46.7	322	2 E82505	conserved hypotet
28	42	46.7	473	2 S70357	forkhead transcrip
29	42	46.7	598	2 F83977	hypothetical prote

30	42	46.7	612	2 T13616	hypothetical prote
31	42	46.7	778	2 A23308	transcription init
32	42	46.7	1071	1 FABYVA	H+-exporting ATPas
33	42	46.7	1998	2 T08822	nonstructural poly
34	41	45.6	264	2 S77408	phosphoribosylamin
35	41	45.6	269	2 AH1970	type 4 prelinin pe
36	41	45.6	302	1 H64127	ribosomal protein
37	41	45.6	335	2 E75183	ATP-binding protei
38	41	45.6	342	2 A10286	pap operon transcr
39	41	45.6	421	2 A12824	NADH dehydrogenase
40	41	45.6	438	2 H97602	probable NADH dehy
41	41	45.6	463	2 F90422	4-hydroxybutyryl-C
42	41	45.6	514	2 T43304	Rad17 protein homo
43	41	45.6	579	2 C71916	aspartate-tRNA lig
44	41	45.6	606	2 T27035	hypothetical prote
45	41	45.6	772	2 S22660	excision repair pr

ALIGNMENTS

RESULT 1

G64131

sporulation protein spoIIIE - Haemophilus influenzae (strain Rd KW20)

C/Species: Haemophilus influenzae

C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004

C/Accession: G64131

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J. D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995

A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A;Reference number: A64000; MUID:95350630; PMID:7542800

A;Accession: G64131

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-529 <TIGR>

A;Cross-references: UNIPROT:P45264; GB:U32833; GB:L42023; NID:G1574432; PIDN:AAC23240.1;

C;Superfamily: Bacillus subtilis DNA translocase spoIIIE

Query Match 55.6%; Score 50; DB 2; Length 529;
Best Local Similarity 53.3%; Pred. No. 4.3;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 ELEGLEPIPNPLRTG 17

Db 286 EAMGMPVPNPIWRLG 300

RESULT 2

JC4767

E3 gene encoding hypothetical 15.0k protein - human adenovirus 35

C/Species: Mastadenovirus h35 (human adenovirus 35)

C/Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004

C/Accession: JC4767

R;Basler, C.F.; Droguett, G.; Horwitz, M.S. Gene 170, 249-254, 1996

A;Title: Sequence of the immunoregulatory early region 3 and flanking sequences of adeno

A;Reference number: JC4765; MUID:96235144; PMID:8666254

A;Accession: JC4767

A;Molecule type: DNA

A;Residues: 1-131 <BAS>

A;Cross-references: UNIPROT:Q67729; GB:U32664; NID:G984529; PIDN:AAA75324.1; PID:G984532

C;Genetics:

A;Gene: E3

C;Superfamily: adenovirus early E3 16K glycoprotein

Query Match 53.3%; Score 48; DB 2; Length 131;

Best Local Similarity 56.2%; Pred. No. 1.8;

Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 2 GELEGLEPIPNPLRTG 17

```
Db      77  GELHGLPTENPWEAG 92
      ||| ||| ||| |||
RESULT 3
AS4265
C/CAAT/enhancer-binding protein - rat
C/Alternate names: C/EBP
C/Species: Rattus norvegicus (Norway rat)
C/Date: 06-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 09-Jul-2004
C/Accession: S06890; A54265
R/Jandschulz, W.H.; Johnson, P.F.; Adashi, E.Y.; Graves, B.J.; McKnight, S.L.
Genes Dev. 2, 786-800, 1988
A/Title: Isolation of a recombinant copy of the gene encoding C/EBP.
A/Reference number: S06890; MUID:9092001; PMID:2850264
A/Accession: S06890
A/Molecule type: DNA
A/Residues: 1-29, RLWLSRGPRAPSTCPGAGR', 55-358 <LA2>
A/Cross-references: UNIPROT:P05554; EMBL:X12752
A/Note: the authors translated the codon AGC for residue 21 as Asp
A/Note: part of this sequence was confirmed by protein sequencing
A/Note: this sequence has been corrected
R/Lincoln, A.J.; Williams, S.C.; Johnson, P.F.
Genes Dev. 8, 1131-1132, 1994
A/Title: A revised sequence of the rat C/EBP gene.
A/Reference number: A54265; MUID:95011606; PMID:7926792
A/Accession: A54265
A/Molecule type: DNA
A/Residues: 1-60 <LIN>
A/Cross-references: GB:X12752
C/Superfamily: CCAAT/enhancer-binding protein alpha
C/Keywords: transcription factor

Query Match      52.2%; Score 47; DB 2; Length 358;
Best Local Similarity 62.5%; Pred. No. 8.4;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      2  GELEGLPIPNPLRTG 17
      ||| ||| ||| |||
Db      247  GSLKGLAGPFDLTG 262
      ||| ||| ||| |||

RESULT 4
C69280
conserved hypothetical protein AF0243 - Archaeoglobus fulgidus
C/Species: Archaeoglobus fulgidus
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C/Accession: C69280
R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A/Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A/Reference number: A69250; MUID:98049343; PMID:9389475
A/Accession: C69280
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-529 <KLE>
A/Cross-references: UNIPROT:O29996; GB:AR001089; GB:AR000782; NID:G2689412; PIDN:AAB9099
C/Superfamily: Pyrococcus abyssi hypothetical protein PAB0059

Query Match      52.2%; Score 47; DB 2; Length 529;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      2  GELEGLPIPNP 12
      ||| ||| |||
Db      222  GEFEGVPRPNP 232
      ||| ||| |||

RESULT 5
AE2706
conserved hypothetical protein Atul054 [imported] - Agrobacterium tumefaciens (strain C58)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C/Accession: AE2706
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AE2706
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-255 <KUR>
A/Cross-references: UNIPROT:Q8UGI3; GB:AE008688; PIDN:AAL42067.1; PID:gl7739446; GSPDB:G
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atul054
A/Map position: circular chromosome

Query Match      51.1%; Score 46; DB 2; Length 255;
Best Local Similarity 42.9%; Pred. No. 8.2;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY      1  KGELEGLPIPNPLL 14
      ||| ||| ||| |||
Db      68  EGSDGIPDPPLV 81
      ||| ||| ||| |||

RESULT 6
E97488
hypothetical protein AGR_C_1946 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C/Accession: E97488
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A/Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: E97488
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-255 <KUR>
A/Cross-references: UNIPROT:Q8UGI3; GB:AE007869; PIDN:AAK86862.1; PID:gl5156078; GSPDB:G
C/Genetics:
A/Gene: AGR_C_1946
A/Map position: circular chromosome

Query Match      51.1%; Score 46; DB 2; Length 255;
Best Local Similarity 42.9%; Pred. No. 8.2;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY      1  KGELEGLPIPNPLL 14
      ||| ||| ||| |||
Db      68  EGSDGIPDPPLV 81
      ||| ||| ||| |||

RESULT 7
AF3320
Flavohemoprotein [imported] - Brucella melitensis (strain 16M)
C/Species: Brucella melitensis
C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 16-Aug-2004
C/Accession: AF3320
R/DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.
; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesc
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A/Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A/Reference number: AD3252; PMID:11756688
A/Accession: AF3320
```

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <KUR>
A:Cross-references: UNIPROT:Q8YI97; GB:AE008917; PIDN:AAL51729.1; PID:G17982465; GSPDB:C
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0548
A:Map position: I
C:Superfamily: Phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferred

Query Match 50.0%; Score 45; DB 2; Length 378;
Best Local Similarity 46.7%; Pred. No. 19;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 ELEGLEPIPNPLRTG 17
| : : : : : |
Db 266 EISAVVPSPPLPETG 280

RESULT 8
S51600
Phosphorylase kinase (EC 2.7.1.38) gamma chain - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 16-Aug-2004
C:Accession: S51600
R:Bahri, S.M.; Chia, W.
Mol. Gen. Genet. 245, 588-597, 1994
A:Title: DPHK-gamma, a putative Drosophila kinase with homology to vertebrate phosphoryl
utants.
A:Reference number: S51600; MUID:95107257; PMID:7808409
A:Accession: S51600
A:Molecule type: DNA
A:Residues: 1-560 <BAH>
A:Cross-references: UNIPROT:Q9YVV7; EMBL:U13014; NID:9531470; PIDN:AAA64560.1; PID:95314
A:Note: the authors translated the codon CAT for residue 528 as Ala and CAC for residue
C:Genetics:
A:Gene: FlyBase:PhK-gamma
A:Cross-references: FlyBase:FBgn0011754
A:Map position: 10
C:Superfamily: protein kinase homology
C:Keywords: phosphotransferase
P:21-291/Domain: protein kinase homology <KIN>

Query Match 50.0%; Score 45; DB 2; Length 560;
Best Local Similarity 72.7%; Pred. No. 30;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 7 LPINPLRTG 17
| : : : : : |
Db 439 LQLENPLLTG 449

RESULT 9
A39283
gamma-glutamyl carboxylase (EC 4.1.1.-) - human
C:Species: Homo sapiens (man)
C>Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 06-Dec-1996
C:Accession: A39283
R:Wu, S.M.; Cheung, W.F.; Frazier, D.; Stafford, D.W.
Science 254, 1634-1636, 1991
A:Title: Cloning and expression of the cDNA for human gamma-glutamyl carboxylase.
A:Reference number: A39283; MUID:92086858; PMID:1749935
A:Accession: A39283
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: mRNA
A:Residues: 1-757 <WUA>
A:Cross-references: GB:M81592
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 50.0%; Score 45; DB 2; Length 757;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KGELEGLPIPNPLRT 16
| : : : : : |
Db 643 EGEVKGGEPTPLVQT 658

RESULT 10
A47439
gamma-glutamyl carboxylase (EC 4.1.1.-) - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C:Accession: A47439; A38939; B39283; C39283
R:Rehemulla, A.; Roth, D.A.; Wasley, L.C.; Kuliopulos, A.; Walsh, C.T.; Furie, B.; Furi
Proc. Natl. Acad. Sci. U.S.A. 90, 4611-4615, 1993
A:Title: In vitro and in vivo functional characterization of bovine vitamin K-dependent
A:Reference number: A47439; MUID:93281629; PMID:8506307
A:Accession: A47439
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-758 <REH>
A:Cross-references: UNIPROT:Q07175; GB:L09726; NID:9289398; PIDN:AAA30425.1; PID:9289399
R:Wu, S.M.; Cheung, W.F.; Frazier, D.; Stafford, D.W.
submitted to GenBank, January 1992
A:Reference number: A38939
A:Accession: A38939
A:Molecule type: mRNA
A:Residues: 96-758 <WUA>
A:Cross-references: GB:M81593; NID:9162744; PIDN:AAA30410.1; PID:9162745
R:Wu, S.M.; Cheung, W.F.; Frazier, D.; Stafford, D.W.
Science 254, 1634-1636, 1991
A:Title: Cloning and expression of the cDNA for human gamma-glutamyl carboxylase.
A:Reference number: A39283; MUID:92086858; PMID:1749935
A:Accession: B39283
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 96-693,695-758 <WU2>
A:Cross-references: GB:M81593
A:Accession: C39283
A:Molecule type: protein
A:Residues: 37-50 <WU3>
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 50.0%; Score 45; DB 2; Length 758;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KGELEGLPIPNPLRT 16
| : : : : : |
Db 643 EGEVKGGEPTPLVQT 658

RESULT 11
S77521
sensory transduction histidine kinase sll1124 - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein sll1124
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C:Accession: S77521
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S77521
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1371 <KAN>
A:Cross-references: UNIPROT:P73337; EMBL:D90905; GB:AB001339; NID:91652360; PIDN:BAAL1736
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 50.0%; Score 45; DB 2; Length 1371;
Best Local Similarity 50.0%; Pred. No. 82;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:47:36 ; Search time 136.87 Seconds
(without alignments)
1189.648 Million cell updates/sec

Title: US-10-622-893A-5
Perfect score: 2530
Sequence: 1 MGSGDDDLALATKSSSL.....LEGIPNPLRTGHHHHH 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1462099 seqs, 344972447 residues

Total number of hits satisfying chosen parameters: 1462099

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA: *
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_PUBCOMB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2530	100.0	472	US-10-622-893A-5	Sequence 5, Appli
2	2329	92.1	437	US-10-622-893A-3	Sequence 3, Appli
3	1153.5	45.6	440	US-10-470-678-9	Sequence 9, Appli
4	1152.5	45.6	440	US-10-470-678-10	Sequence 10, Appli
5	1151.5	45.5	440	US-10-470-678-8	Sequence 8, Appli
6	627	24.8	437	US-10-232-655-3	Sequence 3, Appli
7	600	23.7	437	US-10-232-655-1	Sequence 1, Appli
8	210.5	8.3	139	US-10-767-701-53616	Sequence 53616, A
9	184	7.3	384	US-10-156-761-14478	Sequence 14478, A
10	168	6.6	433	US-10-287-274-412	Sequence 412, App
11	168	6.6	433	US-10-282-122A-42701	Sequence 42701, A
12	166	6.6	387	US-10-829-427-1	Sequence 1, Appli
13	165	6.5	385	US-10-282-122A-68225	Sequence 68225, A

14	159.5	6.3	381	14	US-10-156-761-9328	Sequence 9328, Ap
15	140	5.5	374	15	US-10-282-122A-55515	Sequence 55515, A
16	137.5	5.4	420	15	US-10-425-114-54552	Sequence 54552, A
17	136.5	5.4	372	15	US-10-282-122A-43213	Sequence 43213, A
18	136.5	5.4	383	9	US-09-883-096-5	Sequence 5, Appli
19	135.5	5.4	378	10	US-09-975-719-367	Sequence 367, App
20	127.5	5.0	433	9	US-09-815-242-13745	Sequence 13745, A
21	127	5.0	372	15	US-10-282-122A-75876	Sequence 75876, A
22	127	5.0	414	16	US-10-437-963-177242	Sequence 177242, A
23	126	5.0	370	15	US-10-282-122A-72847	Sequence 72847, A
24	125.5	5.0	570	17	US-10-800-350-386	Sequence 386, App
25	124	4.9	429	15	US-10-425-114-71622	Sequence 71622, A
26	122	4.8	406	17	US-10-485-710-80	Sequence 80, Appl
27	121.5	4.8	669	9	US-09-213-888-25	Sequence 25, Appl
28	121.5	4.8	669	9	US-09-328-877A-25	Sequence 25, Appl
29	121.5	4.8	669	16	US-10-653-497-25	Sequence 25, Appl
30	121.5	4.8	669	16	US-10-653-517-25	Sequence 25, Appl
31	121.5	4.8	669	16	US-10-653-496A-25	Sequence 25, Appl
32	121.5	4.8	669	17	US-10-653-676A-25	Sequence 25, Appl
33	120.5	4.8	405	14	US-10-156-761-9148	Sequence 9148, Ap
34	120	4.7	818	14	US-10-156-761-14479	Sequence 14479, A
35	119.5	4.7	391	15	US-10-282-122A-69762	Sequence 69762, A
36	115.5	4.6	568	10	US-09-953-348-135	Sequence 135, App
37	115.5	4.6	568	14	US-10-267-255-135	Sequence 135, App
38	115	4.5	563	15	US-10-415-302-8	Sequence 8, Appli
39	115	4.5	563	15	US-10-415-302-18	Sequence 18, Appl
40	113.5	4.5	395	9	US-09-873-880-30	Sequence 30, Appl
41	113.5	4.5	395	15	US-10-424-599-268694	Sequence 268694, A
42	113.5	4.5	542	16	US-10-437-963-117091	Sequence 117091, A
43	113.5	4.5	632	15	US-10-282-122A-68951	Sequence 68951, A
44	113.5	4.5	832	15	US-10-282-122A-50546	Sequence 50546, A
45	112.5	4.4	317	15	US-10-282-122A-61710	Sequence 61710, A

ALIGNMENTS

RESULT 1
US-10-622-893A-5
; Sequence 5, Application US/10622893A
; Publication No. US20050014935A1
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Sheng
; APPLICANT: Datta, Abhijit
; APPLICANT: Wang, Yuping
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DETERMINATION OF GLYCATED PROTEINS
; FILE REFERENCE: 466992001300
; CURRENT APPLICATION NUMBER: US/10/622,893A
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric protein
US-10-622-893A-5

Query Match	100.0%	Score	2530	DB	17	Length	472
Best Local Similarity	100.0%	Pred. No.	7.6e-237				
Matches	472	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MGSGDDDLALATKSSSLIVGAGTGTSTALHARRGVTNTVLDPPVPSAISAGN	60				
Db	1	MGSGDDDLALATKSSSLIVGAGTGTSTALHARRGVTNTVLDPPVPSAISAGN	60				
Qy	61	DVKNVSSGQYNNKDKIEVNEILAEAFNGKNDPLFKPYHDTGLLMSACSQEGDLRL	120				
Db	61	DVKNVSSGQYNNKDKIEVNEILAEAFNGKNDPLFKPYHDTGLLMSACSQEGDLRL	120				
Qy	121	GVRVRPEDPNLVELTRPEQFKLAPGVLQDPPGKGYFARSGAGWAHARNALVAAR	180				

Db 121 GVRVREGEDNVLVELTRPEQFRKLAPEGVLOGDFPGWKGYFARSGAGWAHARNALVAAAR 180
QY 181 EAQRMGVKFTGTPQGRVVTLIFENNNDVGAVTGDKIWAERTFLCAGASAGQFLDFKN 240
Db 181 EAQRMGVKFTGTPQGRVVTLIFENNNDVGAVTGDKIWAERTFLCAGASAGQFLDFKN 240
QY 241 QURPTAATLVHIALKPEERALKYKNIPIVFNIERGFFPEDEERGEIKICDEHPGYTNMVQ 300
Db 241 QURPTAATLVHIALKPEERALKYKNIPIVFNIERGFFPEDEERGEIKICDEHPGYTNMVQ 300
QY 301 SADGTWMSIPFEKTQIPKEAETRVALLKETMPQADRPFSFARICWCADTANREFLIDR 360
Db 301 SADGTWMSIPFEKTQIPKEAETRVALLKETMPQADRPFSFARICWCADTANREFLIDR 360
QY 361 HPQYHSLVLCGASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANRNWRDYL 420
Db 361 HPQYHSLVLCGASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANRNWRDYL 420
QY 421 GRFGGPNRVMDFDHVKWNTNVQYRDISKLGELGPIPNPLLRTHGHHHH 472
Db 421 GRFGGPNRVMDFDHVKWNTNVQYRDISKLGELGPIPNPLLRTHGHHHH 472

RESULT 2

US-10-622-893A-3
; Sequence 3, Application US/10622893A
; Publication No. US20050014935A1
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Sheng
; APPLICANT: Datta, Abhijit
; APPLICANT: Wang, Yuping
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DETERMINATION OF GLYCATED PROTEINS
; FILE REFERENCE: 466992001300
; CURRENT APPLICATION NUMBER: US/10/622, 893A
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 40%-100% identity to the amadoriase
US-10-622-893A-3

Query Match 92.1%; Score 2329; DB 17; Length 437;
Best Local Similarity 100.0%; Pred. No. 2.4e-217;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 AVTKSSLLIVGAGTGTSTALHLARRGYTNVTLDPPVPSPISAGNDVNVKVISSGOYS 72
Db 1 AVTKSSLLIVGAGTGTSTALHLARRGYTNVTLDPPVPSPISAGNDVNVKVISSGOYS 60
QY 73 NNKDEIWEIILAEAEFNGKNDPLFKPYHYDHTGLLMSACSQGLDRLGVVRPGEEDPNL 132
Db 61 NNKDEIWEIILAEAEFNGKNDPLFKPYHYDHTGLLMSACSQGLDRLGVVRPGEEDPNL 120
QY 133 VELTRPEQFRKLAPEGVLOGDFPGWKGYFARSGAGWAHARNALVAAAREAQRGVKPVVTG 192
Db 121 VELTRPEQFRKLAPEGVLOGDFPGWKGYFARSGAGWAHARNALVAAAREAQRGVKPVVTG 180
QY 193 TPQGRVVTLIFENNNDVGAVTGDKIWAERTFLCAGASAGQFLDFKNLRPTAATLVHI 252
Db 181 TPQGRVVTLIFENNNDVGAVTGDKIWAERTFLCAGASAGQFLDFKNLRPTAATLVHI 240
QY 253 ALKPEERALKYKNIPIVFNIERGFFPEDEERGEIKICDEHPGYTNMVQSDGTWMSIPFE 312
Db 241 ALKPEERALKYKNIPIVFNIERGFFPEDEERGEIKICDEHPGYTNMVQSDGTWMSIPFE 300
QY 313 KTOIPKEAETRVALLKETMPQADRPFSFARICWCADTANREFLIDRHPQYHSLVLCG 372

Db 301 KTOIPKEAETRVALLKETMPQADRPFSFARICWCADTANREFLIDRHPQYHSLVLCG 360
QY 373 ASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANRNWRDYLGRFGGPNRVMDP 432
Db 361 ASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANRNWRDYLGRFGGPNRVMDP 420
QY 433 HDVKWNTNVQYRDISKL 449
Db 421 HDVKWNTNVQYRDISKL 437
RESULT 3
US-10-470-678-9
; Sequence 9, Application US/10470678
; Publication No. US20050101711A1
; GENERAL INFORMATION:
; APPLICANT: KOUZUMA, Takuji et al.
; TITLE OF INVENTION: COMPOSITION FOR ASSAYING GLYCATED PROTEINS
; FILE REFERENCE: 1516-0121P
; CURRENT APPLICATION NUMBER: US/10/470, 678
; PRIOR FILING DATE: 2003-07-30
; PRIOR APPLICATION NUMBER: JP 2001/ 22953
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: JP 2001/ 39796
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: JP 2001/ 24002
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic protein sequence of SEQ ID NO: 3
; NAME/KEY: MUTAGEN
; LOCATION: (372)..(372)
US-10-470-678-9

Query Match 45.6%; Score 1153.5; DB 17; Length 440;
Best Local Similarity 50.9%; Pred. No. 5.6e-103;
Matches 223; Conservative 75; Mismatches 133; Indels 7; Gaps 5;
QY 11 ALAVTKSSLLIVGAGTGTSTALHLARRGYTNVTLDPPVPSPISAGNDVNVKVIS--S 68
Db 1 ASTLTQSQQLIVGGTGCSTALHLARRGYTNVTLDVNRIPSPISAGHDVNVKLAGRLS 60
QY 69 GOYSNNKDEIWEIILAEAEFNGKNDPLFKPYHYDHTGLLMSACSQGLDRLGVVRPGE 128
Db 61 TADSKGDDSDSIWKALSYAAQGLHDPVFPQFCHNTGVSIVAGSTPKSIKQL-VEDEIGD 119
QY 129 D-PNLVELTRPEQFRKLAPEGVLOGDFPGWKGYFARSGAGWAHARNALVAAAREAQRMGV 187
Db 120 DIDQYTPLTAEDFRKTMPEGLITGFPGWKGYKFTGSGWVHARKAMKRAAFESRELGV 179
QY 188 KFTVTGTPQGRVVTLIFENNNDVGAVTGDKIWAERTFLCAGASAGQFLDFKNQLRPTAW 247
Db 180 KFTGSPGKVESLIIFEDGDVGRGAKTADGKEHADRTILSAGASAEFFLDQFNQIOPTAW 239
QY 248 TLVHIALKPEERALKYKNIPIVFNIERGFFPEDEERGEIKICDEHPGYTNMVQSDGTM- 306
Db 240 TLGHIQMTPEETKLYKNLPPLEFNINQGFMEPPEDDLHQLKMCDEHPGYNWVKPGSKYP 299
QY 307 MSTPFEKTQIPKEAETRVALLKETMPQADRPFSFARICWCADTANREFLIDRHPQYHS 366
Db 300 QSIFFAKHQVPTAEARMKQFLKDIIMPQADRLVHARICWCADTQORMFLIYHPHPS 359
QY 367 LVLCGASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNPDIANRNWRDYLGRFGGP 426
Db 360 LVIASGDCGTGYMHIIISIGKFIISDCMEGTLEERFAKFWWRPEKFTFEWKGDLDRFGAD 419
QY 427 NRVMDP--HDVKWNTNVQ 442


```
; CURRENT APPLICATION NUMBER: US/10/232,655
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Eupenicillium terrenum
US-10-232-655-3

Query Match      24.8%; Score 627; DB 14; Length 437;
Best Local Similarity 34.1%; Pred. No. 9.2e-52;
Matches 153; Conservative 82; Mismatches 174; Indels 40; Gaps 14;

QY 12 LAVTKSSLLIV--GAGTGTSTALHLARRGYT--NVTVLDRYPVPSAISAGNDVNKVIS 67
Db 1 MAHSRASTKVVVVGCGGTIGSSSTALHLIRSGYTPSNITVLDVYKTESLOSAGHDLNKM-- 59
QY 68 SGQYSNNKDEIEVNEILAEAFNGKNDPLFKPYHYDHTGLLMSACSQEGLDRLGLVVRP- 126
Db 60 -GIRLRNGPDLQ----LSLESIDMWQNDLDFKPFHQVGMIDCSSKEGIEENLRKYQTL 114
QY 127 -----GEDPNLVELTRPEQFKLAPEGVLQGD--PGWKGYFARGAGWAHARNALVAAARE 181
Db 115 LDAGIGLEKTNVWLESEDEILAKAP-NPTREQVKGWKLGFCTDG-GWLAATAKAINAIGIF 172
QY 182 AQRMGVKFVTGTPQGRVVTILP--ENNVDKGVATGDKTWRABRTFLCAGASAGQFLDFK 239
Db 173 LQDKGVKFGF-GAGTFQFPFADGKTCIGLETTDGTGKFADKVVLAAGAWSPTLVLDLE 231
QY 240 NQLRPATWLTVHIALKPEERALKYNIPIFNIERGFFFEDEERGEIKICDEHPGYTNM- 298
Db 232 DQCVSKAWVFAHQLTPKEADAYKNVPVYDGYGVGFFEPN-EYGVIKVCDEPFGFSRFK 290
QY 299 -----VOSAGTMMISPEKTOIPKE-----AETRVRALLKETMPQLADRPFSFARICWCA 349
Db 291 LHQPYGAASGPKMISVPSRSHAKHPTDTPDASEVTIRKAIARFLPEFKDELFRNTMCWCT 350
QY 350 DTANREFLIDRHPQYHSLVLCGASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNP 409
Db 351 DTADANLLICEHPKKNFILTGDGSHSFKLLENKGVHVELLEGSLSQEMAGAWRRPG 410
QY 410 IAAANRNRDTLGRFGGPNRVMDFHVKW 438
Db 411 GDALRS-----RRGAPAK--DLAEMPGW 431

RESULT 7
US-10-232-655-1
; Sequence 1, Application US/10232655
; Publication No. US20030157593A1
; GENERAL INFORMATION:
; APPLICANT: KUROSAWA, KEIKO
; APPLICANT: HIROKAWA, KOZO
; APPLICANT: KAJIYAMA, NAOKI
; TITLE OF INVENTION: No. US20030157593A1e1 fructosyl peptide oxidase
; FILE REFERENCE: 227590U0
; CURRENT APPLICATION NUMBER: US/10/232,655
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 1
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Coniochaeta sp.
US-10-232-655-1

Query Match      23.7%; Score 600; DB 14; Length 437;
Best Local Similarity 32.7%; Pred. No. 3.9e-49;
Matches 141; Conservative 86; Mismatches 162; Indels 42; Gaps 13;

QY 20 LLIVGAGTGTSTALHLARRGY--TNVTVLDRYPVPSAISAGNDVNKVISGQYSNNKDE 77
Db 11 IVVGGGTIGSSSTALHLIRSGYAPANITVLDTFEIPSAQSAGHDLNKM--GIRLRNKVD 68
QY 78 IEVNEILAEAFNGKNDPLFKPYHYDHTGLLMSACSQEGLDRLGLVVRPGEEDPNL-VELT 136
Db 69 LQ-----MSLEARQMWKEDLFPQFFHNTGDMCEHTPEGIEDLKKQYQALHDAGAGLEKT 124
QY 137 R---PEQFKLAPEGVLQGD--PGWKGYFARGAGWAHARNALVAAAREAQRMGVKEVTG 192
Db 125 HAWLDNDEILSKMPLLRDQIQGWKAIWSQDG-GWLAATAKAINAIGQFLKERGVKFGFG 183
QY 193 TPQGRVVTILFENNVDK--GAVTGDGKIWRABRTFLCAGASAGQFLDFKNQLRPATWTLV 250
Db 184 -GAGSPKQFLFDEGTCIGVETADGTKYADKVVLAAGAWSPTLVLDLEQCCSKAWVYA 242
QY 251 HIALKPEERALKYNIPIFNIERGFFFEDEERGEIKICDEHPGYTNMVO-----SADGT 305
Db 243 HIQLTFEEAAEYKGVVPVYNGSFGFFEPN-EFGVLKVCDEPFGFSRFEKHQPYGAPSPK 301
QY 306 MMSIPEKTOIPKE-----AETRVRALLKETMPQLADRPFSFARICWCADTANREFLIDR 360
Db 302 HISVPSRSHAKHPTDTPDASEVSIKKAIAITFLPRFQDKELFNRLALCWCCTDTADAALIMCE 361
QY 361 HPQYHSLVLCGASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKWNP----- 408
Db 362 HPKWKNFILATGDGSHSFKLLENKGVHVELLEGRLPEEMAYQWRWRPGDALKSRRAAP 421
QY 409 --DIAANRNR 417
Db 422 PKDLADMPGWK 432

RESULT 8
US-10-767-701-53616
; Sequence 53616, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 53616
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 13394327.pap
US-10-767-701-53616

Query Match      8.3%; Score 210.5; DB 16; Length 139;
Best Local Similarity 33.6%; Pred. No. 5.3e-12;
Matches 44; Conservative 21; Mismatches 39; Indels 27; Gaps 3;

QY 315 QIPKEAETRVRALLKETMPQLADRPFSFARICWCADTANREFLIDRHPQYHSLVLCGAS 374
Db 7 EIPREGQDVCREFLRCIPEFGNRPWTHTRICWTTDPTGWDWIDYHPKYKGLFVATGGS 66
```


APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 42701
LENGTH: 433
TYPE: PRT
ORGANISM: Escherichia coli
US-10-282-122A-42701

Query Match 6.6%; Score 168; DB 15; Length 433;
Best Local Similarity 23.0%; Pred. No. 4.1e-07;
Matches 120; Conservative 59; Mismatches 176; Indels 166; Gaps 28;
QY 1 MGSGGDDDLALAVTKSSLLIIVAGTGTSTALHARGYNTVTLDPYPVPSAISAGN 60
DB 7 LGHDMEDD-----CDIIITIGAGTAGTACALCARAG-LSVLLERAEIPG----- 50
QY 61 DVNKVISGGQYSNNKDBIEVNEILAEAFNGWKNDPLFKPYHDTGLLMSACSOEGLDRL 120
DB 51 --SKNLSGGRL-----YTHALAE-----LLPQHLTAPLERRITHESLSLL 89
QY 121 ---GVRVRFGEDPNLVELTRPEQFRKLAPEG-----VLQGDFFGKGYFARSGAGWAHARN 173
DB 90 TPDGVTT-----FSLQPGGESWSVLARFDFW----- 117
QY 174 ALVAAREARMGVKFTVTPQGRVVTLLIFENNVDKAVTGDGKIWAERTFLCAGASAG 233
DB 118 -LVA---EAEKEGVECI---PGATVDALYBENGRCVIGCD--DILRARYVVLAEGA--- 166
QY 234 QFLDFKNQLRPTAWTLVHIALKPEERALKYKNIPIVFN-----TERGFFPEDEERGEI-- 286
DB 167 -----NSVLAERHGLV---TRPAGEAMALGIKEVLSLETSALAEERPHLENNEGAALLFS 217
QY 287 -KICDBHPG---YTNVQASDG-----TWMSIPFEXTQIPKEATRVRLALKKET-- 331
DB 218 GRICDDLPGGAFLYTNQOTLSLGLVCPVLSLTSQSRVPASSELLTRFAHPAVRPLIKNTES 277
QY 332 -----MPQ--LADRPFSFARICW-----CADTANREFLIDRHPQYHSLVLCGGA 373
DB 278 LEYGAHLVPEGLHSMFVQYAGNGWLLVGLDALRSCVNT-----GI 317

QY 374 SGRGFKY----LPSIGNLIIVDAMEGKVPQKIHKLKWNPDIAANRWRDTLGRFGG-PNR 428
DB 318 SVRGDMALTGAQAAATLISACQHEPQNLPL--YHNVERSLLM-DVLQRYQHPAL 374
QY 429 VMDFDHVKWNTVQYVRDISKLKGELEGLPIPNPLLTGHHH 469
DB 375 LQRPGWYRTWPALM-ODISRLMDQGDKEVP-PLRQLFWHH 413

RESULT 12
US-10-829-427-1
Sequence 1, Application US/10829427
Publication No. US20050026265A1
GENERAL INFORMATION:
APPLICANT: FURUKAWA, KEISUKE
APPLICANT: KAJIYAMA, NAOKI
TITLE OF INVENTION: MODIFIED SARCOSEINE OXIDASES, GENES AND RECOMBINANT DNAS THEREOF,
FILE REFERENCE: 252202US0
CURRENT APPLICATION NUMBER: US/10/829,427
CURRENT FILING DATE: 2004-04-22
PRIOR APPLICATION NUMBER: JP 2003-121533
PRIOR FILING DATE: 2003-04-25
PRIOR APPLICATION NUMBER: JP 2003-396807
PRIOR FILING DATE: 2003-11-27
PRIOR APPLICATION NUMBER: JP 2004-116345
PRIOR FILING DATE: 2004-04-12
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.3
SEQ ID NO 1
LENGTH: 387
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
US-10-829-427-1

Query Match 6.6%; Score 166; DB 17; Length 387;
Best Local Similarity 21.4%; Pred. No. 5.4e-07;
Matches 89; Conservative 68; Mismatches 173; Indels 86; Gaps 20;
QY 20 LLIIVGAGTGTSTALHARGYNTVTLDPYPVPSAISAGNDVKNKVISSQYSNNKDEIE 79
DB 7 VIVVGAGSGMAAGYYLAKQGVKTLV-DAFDPPHTEGSHHGDTIRIHA-YGEGRKYP 64
QY 80 VNEILAEAFNGWKNDPLFKPYHDTGLL-----MSACSOEGL-----D 118
DB 65 F-ALRAQELWYLENETHNK-IFTKTGVLVFGPKGESDFVAETMEAAAEHSLTVDLLEGD 122
QY 119 RLGVVRFGEDPNLVELTRPEQFRKLAPEGVLQGDFFGKGYFARSGAGWAHARNALVAA 178
DB 123 EINTR-WPG-----ITVPENYNAI-----FEPNSGVLFENCIRSY 157
QY 179 AREAARMGVKFTVTPQGRVVTLLIFENNVDKAVTGDGKIWAERTFLCAGASAGFLDF 238
DB 158 RELAVAKGAKILTYT--RVEDFEVSQDVK-IQTANGS-YTADKLIVSGAMNSKLSK 212
QY 239 KNQLRPTAWTLVHIALKPEERALKYKNIPIVFN-----EERALKYKNIPIVFNTERGFFPEDEERGE 285
DB 213 LN-----LDIPLQYRVQVVGFPDSDNEAKYSNDVGYPAFMVPEVPKGIYGFPSFGG- 262
QY 286 IKICDBHPGVTNMQASDGTWMSIPFEKTQIPKEATRVRLALKETMPQLADRPFSFARI 345
DB 263 ---CGLUKIGYHTYVGGQIDPDTINREFGAYQ---EDESRLRDFLEKYMPE-ANGELKRGAV 315
QY 346 CWCADTANREFLIDRHPQYHSLVLCGASGRGFKYLFISIGNLIIVD-AMEGKVPQKI 400
DB 316 CMYTKPTDHHFVTDTHPEHNSVFAAGFSGHGFSSVVGVLSQLATTGKTEHDI 371

RESULT 13
US-10-282-122A-68225

```
Sequence 68225, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsuo
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zvekind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 68225
LENGTH: 385
TYPE: PRT
ORGANISM: Pseudomonas putida
US-10-282-122A-68225

Query Match 6.58; Score 165; DB 15; Length 385;
Best Local Similarity 23.5%; Pred. No. 6.7e-07;
Matches 101; Conservative 53; Mismatches 161; Indels 114; Gaps 22;

QY 16 KSSLLIVGAGTGTSTALHARRGYNTVLDPPVPSPSAISAGNDVKNVSS-----68
Db 6 KSTAVVVGUGVGSATLWFLAQ--QQDVL-----CLEGAPINLOGSSYGSRIF 55
QY 69 -GOYSNNKDBIEVNEILAEAFNGWK--NDPLFKPYVYDHTGLMSACSQBGLDRLGVVR 125
Db 56 ROAYWEGSDYLS---LLAEADLGWRELQATSHRPLHYSG-----GLFIG 97
QY 126 PGEDPNLVELTRPQRFKLAPEGVLQDFFGKGYFARSGAGWAHARNALVAAAREA---182
Db 98 P-----IRSGVVGSAAS-----AKAG-GIAHOR--LTAAEVEARFS 131
QY 183 -----QRWGVKFTVTG-----PQGRVVTILPENNDVKAV---TG 214
Db 132 VFRADENMEAVFQGAFTIAADARLQMLNQAVAHGAQMFQSHVQDITRVESEFLLRLS 191
QY 215 DGKIMRAERTFLCAGAS-AGQFL-DFKNQLRPTAHTLVHIALKPEERALKYIPV-IFNI 271
Db 192 DGQSVLAQKVVLATGAGLAGSLIPDLGLRPSRVPYIWCFAFKSGAEQLFANFPATLYEL 251
QY 272 ERG--PFFEPDEERGE--IKICDEHPGYTNMVOSA-DGTMMSIPFEKTQIPKEAETRVRA 326

; Sequence 9328, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9328
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9328

Query Match 6.3%; Score 159.5; DB 14; Length 381;
Best Local Similarity 23.3%; Pred. No. 2.2e-06;
Matches 96; Conservative 53; Mismatches 172; Indels 91; Gaps 17;

QY 18 SLLIVGAGTGTSTALHARRGYNTVLDPPVPSPSAISAGNDVKNVSSGQYNNKDE 77
Db 5 ADVVIGGVGVTSTIAYHLARAGVVDVLRDELASGSTR-----AAGGVRAQPSDE 58
QY 78 IEVN-EILAEAFNGWKNDPLFKPYVYDHTG--LLMSACSQBGLDRLGVVRP--GEDPNL 132
Db 59 LNIRLGARSLAEAFRGEGEPGQDIGLHRVGYLFLSTPEEVAAFETGVLRLQNSLGVPSRM 118
QY 133 VELTRPQRFKLAPE-----EGVLQDFFGKGYFARSGAGWAHARNALVAAAREAQRMGVK 188
Db 119 LE---PAEARRLSPLITDGLLAAAFSPDDGCHCTPESVVHGYA-----AAR---RHGAK 167
QY 189 F-----VTGTPOGRVVTILFENNNDVKAVTGDGKIWAERTFLCA-----GASAGQF 235
Db 168 VLRHCEVTG-----IETRGDDITAVLTRGRRI--GTGTVICAAGAWSKAVGAMVGD 217
QY 236 LDFKNQLRPTAHTLVHIALKPEERALKYIPVIFNTERGFFFPDEERGEIKICDEHPGY 295
Db 218 LPVRPLRRQIAVTEPVTVGLPP-----GLPMTIDFTSSLYFH-----TEGPGL 259
QY 296 TNMVQASDGTMSIPPEKTQIPKEAETRVRLIKETMPOLADRPFSFARICWCADTANRE 355
Db 260 L-----VGMSDPDDETTGFTSTHTRWIPRLCEAMERRAPPLLLDR-----TGGWA 305
QY 356 FLIDRHPQYHSLV-----LGCAGSRGFKYLPISIGNLVDMEGKVP 397
Db 306 GLVEITPDHNLIGEAGSCSRFLYATGFGSHGFLQGPVAGVEVIRDLVLRVP 357

RESULT 14
US-10-156-761-9328
; Sequence 9328, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9328
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9328

; Sequence 68225, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsuo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68225
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Pseudomonas putida
US-10-282-122A-68225
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:28:55 ; Search time 103.659 Seconds
(without alignments)
2331.700 Million cell updates/sec

Title: US-10-622-893A-5

Perfect score: 2530

Sequence: 1 MGSGGDDDLAVTKSSSL.....LEGPIPNPLRTGHHHHH 472

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03:*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2331	92.1	438	2 P78573	P78573 aspergillus
2	1982.5	78.4	437	2 Q92204	Q92204 aspergillus
3	1242.5	49.1	445	2 Q42629	Q42629 aspergillus
4	848	33.5	426	2 Q6BZB0	Q6BZB0 debaryomyce
5	846.5	33.5	418	2 Q6BZA7	Q6BZA7 debaryomyce
6	660	26.1	437	2 Q92261	Q92261 penicillium
7	629	24.9	438	2 Q96UT4	Q96UT4 emericella
8	627	24.8	437	2 Q76SA9	Q76SA9 eupenicilli
9	600	23.7	437	2 Q76SA8	Q76SA8 coniochaeta
10	488.5	19.3	412	2 Q43029	Q43029 schizosacch
11	465	18.4	433	2 Q9UTM9	Q9UTM9 schizosacch
12	426.5	16.9	502	2 Q7S638	Q7S638 neurospora
13	386	15.3	433	2 Q6CL10	Q6CL10 kluyveromyc
14	378.5	15.0	486	2 Q6BN43	Q6BN43 debaryomyce
15	351	13.9	446	2 Q6BN42	Q6BN42 debaryomyce
16	346	13.7	435	2 Q9P976	Q9P976 cylindrocarr
17	192	7.6	390	1 MSOX_BACB0	P40859 bacillus sp
18	185	7.3	387	2 Q6ITC6	Q6ITC6 bacillus ce
19	183	7.2	371	2 Q737A8	Q737A8 bacillus ce
20	181.5	7.2	440	2 Q9V9B5	Q9V9B5 drosophila
21	181	7.2	384	1 MSOX_STRAW	Q827H4 streptomyce
22	181	7.2	388	1 MSOX_STRSB	P40854 streptomyce
23	177.5	7.0	390	1 MSOX_HUMAN	Q9P029 homo sapien
24	177.5	7.0	390	2 Q6IAJ9	Q6IAJ9 homo sapien
25	175	6.9	371	2 Q63AM4	Q63AM4 bacillus ce
26	174	6.9	371	2 Q81PT2	Q81PT2 bacillus an
27	172.5	6.8	395	2 Q92NL2	Q92NL2 rhizobium m
28	172	6.8	367	2 Q982V8	Q982V8 rhizobium l
29	170.5	6.7	773	2 Q66IN9	Q66IN9 xenopus lae
30	170	6.7	371	2 Q81CN0	Q81CN0 bacillus ce
31	170	6.7	371	2 Q6H111	Q6H111 bacillus th

32 170 6.7 399 2 Q95U69 Q95U69 drosophila
33 169.5 6.7 390 1 SOX_RABIT P79371 oryctolagus
34 169 6.7 435 2 Q8XTY8 Q8XTY8 raistonia s
35 166.5 6.6 423 1 YGCN_ECOLI Q46904 escherichia
36 166 6.6 388 1 MSOX_ARTST P40873 arthrobacte
37 165 6.5 382 2 Q88GE9 Q88GE9 pseudomonas
38 162.5 6.4 390 2 Q8N6Z6 Q8N6Z6 homo sapien
39 160 6.3 390 1 SOX_MOUSE Q82826 mus musculu
40 159.5 6.3 381 2 Q82M71 Q82M71 streptomyce
41 159.5 6.3 410 2 Q89M92 Q89M92 bradyrhizob
42 157 6.2 433 2 Q8FEI5 Q8FEI5 escherichia
43 155 6.1 331 2 Q6R5L2 Q6R5L2 bacillus sp
44 155 6.1 375 2 Q8EMP0 Q8EMP0 oceanobacil
45 155 6.1 387 1 MSOX_BACSN P23342 bacillus sp

ALIGNMENTS

RESULT 1

ID P78573 PRELIMINARY; PRT; 438 AA.
AC P78573;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructosyl amine: oxygen oxidoreductase.
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97284723; PubMed=9139700; DOI=10.1074/jbc.272.19.12505;
RA Takahashi M., Pischetsrieder M., Monnier V.M.;
RT "Molecular cloning and expression of amadoriase isoenzyme (fructosyl
RT amine:oxygen oxidoreductase, EC 1.5.3) from Aspergillus fumigatus.";
RL J. Biol. Chem. 272:12505-12507(1997).
DR EMBL; U82830; AAC49711.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR006076; Fad Oxred.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR000594; Thif_domain.
DR Pfam; PF01266; DAO; 1.
SQ SEQUENCE 438 AA; 48931 MW; FBC8A3ESD89D02AE CRC64;

Query Match 92.1%; Score 2331; DB 2; Length 438;
Best Local Similarity 99.8%; Pred. No. 9.5e-171;
Matches 437; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LAVTKSSLLIVGAGTGTSTALHLARRGYTNVTLPYPVPSAISAGNDVNVKVISGQY 71
Db 1 MAVTKSSLLIVGAGTGTSTALHLARRGYTNVTLPYPVPSAISAGNDVNVKVISGQY 60
Qy 72 SNKDEIYVNEIIAEAFNCFWKNDFPKPYHTDTGLMSACSQEGLDRLGVVRPGEDPN 131
Db 61 SNKDEIYVNEIIAEAFNCFWKNDFPKPYHTDTGLMSACSQEGLDRLGVVRPGEDPN 120
Qy 132 LVELTRPEQFRKLAPEGVLQDPPGKGYFARSAGAHARNALVAAREAQRMGVKFVT 191
Db 121 LVELTRPEQFRKLAPEGVLQDPPGKGYFARSAGAHARNALVAAREAQRMGVKFVT 180
Qy 192 GTPQGRVVTLLIFENNVDKAVTGDGKIWAERTFLCAGASAGQFLDFKNQLRPTAWTLVH 251
Db 181 GTPQGRVVTLLIFENNVDKAVTGDGKIWAERTFLCAGASAGQFLDFKNQLRPTAWTLVH 240
Qy 252 IALKPEERALLYKNIPIVFNTERGFFPEPDEBERGIKICDEHPGYTNMVQSDGTMMSIPF 311
Db 241 IALKPEERALLYKNIPIVFNTERGFFPEPDEBERGIKICDEHPGYTNMVQSDGTMMSIPF 300
Qy 312 EKTQIPKEAETRVALLKETMPQLADRFPSFARICWCADTANREFLIDRHPQYHSLVIGC 371
Db 301 EKTQIPKEAETRVALLKETMPQLADRFPSFARICWCADTANREFLIDRHPQYHSLVIGC 360


```
OX NCBI_TaxID=284592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cattolico L., Confaniolieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Kozul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Sennene D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382133; CAG84414.1; -.
DR InterPro; IPR006076; Fad_oxred.
DR InterPro; IPR00205; NAD_BS.
DR Pfam; PF01266; DAO; 1.
SQ SEQUENCE 426 AA; 48297 MW; 7A5A382DF1F72293 CRC64;

Query Match 33.5%; Score 848; DB 2; Length 426;
Best Local Similarity 41.2%; Pred. No. 1.1e-56;
Matches 181; Conservative 71; Mismatches 153; Indels 34; Gaps 10;

QY 16 KSSSLIVAGTGTSTALHLARGYTNVTLDPYVPVPSAISAGNDVNVKVISSQY---Y 71
DB 3 KPGKILIGAGTGLSTALHLRQKQKDVLPVDPYVPVPSAISAGNDVNVKVIQTTSDDDFY 62

QY 72 SNKKDEIVNEIIAEEAFNGKNDPLFKPYHYDHTGLLMSACSOEGLDRLGVVR--PGED 129
DB 63 SK-----LAEALENWRDNVFNKAFETGIYATGKEQESIDRYEYLLGRK 112

QY 130 PNLVELTRPQFRKLAP--EGV--LQGDFFGWKGYFARSAGAGWAHARNALVAAREAQRM 185
DB 113 DKVVKLNSVDYEKYVPNKGKLSYPNKFQWYGYQKNCGWAFARLALENCVCBECKL 172

QY 186 GVKFVTGTPQGRVVTLPF--ENNVDKGAVTGDGKIWAERTFLCAGASAGOFDFKQQLRP 244
DB 173 GAKFVIDSAB-----ELFSDGACVGVHTSGNIIIEADRTIICAGANSFKFLNFEQQLA 228

QY 245 TAWTLVHIALKPERALYKNIPVFNIERGFFPEPDEERGEIKICDEHPGYTNVQVSADG 304
DB 229 KCTYLGHILKLTDEAALLKGPVVLNDLGGFVEPD--LNNEIKFCNEFPFYVNVNED-- 285

QY 305 TWMSIPEKTOIPKEAETRVRALLKETMPQLADRPFSFARICWCAPTANREFLIDRHPQY 364
DB 286 ---SVPSFKDSIPKEAEDOMRFLRQVFPFAEREFSLRICWCTDTPDRHFLICEHPGH 342

QY 365 HSLVLGCGAGRGFKYLPSTGNLIVDM---BKVPOKIHILKNWPDIAANRNRDITLG 421
DB 343 KNLVLGTDGSGQGFYKPNYNGKYSIQVALGENSLDKDKKELWRWRPDMGKRLDKLQ 402

QY 422 RFGGPNRVMDFHDVKEWTN 440
DB 403 RYGSNEVKDLKNVKQWSN 421

RESULT 5
Q6BZA7 PRELIMINARY; PRT; 418 AA.
ID Q6BZA7
AC Q6BZA7;
```

```
QY 428 RVWDFHVKW 438
Db 407 HVKDLKIDW 417

RESULT 6
Q92261 PRELIMINARY; PRT; 437 AA.
AC Q9261;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructosyl amino acid oxidase.
GN Names: faoP;
OS Penicillium janthinellum (Penicillium vitale).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=5079;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97175014; PubMed=9022674;
RA Yoshida N., Sakai Y., Isogai A., Fukuya H., Yagi M., Tani Y., Kato N.;
RT Primary structures of fungal fructosyl amino acid oxidases important
RT for measurement of glycosylated proteins.;
RL Eur. J. Biochem. 242:499-505(1996).
DR EMBL; Y09021; CAA70219.1; -.
DR Pfam; PF01266; DAO; 1.
SQ SEQUENCE 437 AA; 48231 MW; E681B8702DE4A574 CRC64;

Query Match 26.1%; Score 660; DB 2; Length 437;
Best Local Similarity 33.6%; Pred. No. 3.3e-42;
Matches 156; Conservative 75; Mismatches 157; Indels 76; Gaps 14;

QY 16 KSSLLIV-CAGTGTGTSTLHLARRGT--NVTLDVPVPSPSAISAGNDVNVKVISSGQYS 72
Db 6 ESTKIVVGGGTGSGSTLHLIRSGYTPSNITVLDVPPIPSQAGYDLNKIMSI--RL 63

QY 73 NNKDEIVNEILAEAFNGWKNDPLFKPYVYHDTGLLMSACSQGLDRLGVRRPQGDPNL 132
Db 64 RUGPDQ----LSLEALDMWKNDPLFKPFHNVGMIDCSQSGIASLRKHQDLIDANI 119

QY 133 -----VELTRPQFKLAPEGVLQDFFPGWKGYFARSGAGWAHARNALVAAAREAQRMG 186
Db 120 GLEKTNWLESEDDILAKAPH-FTREQIKGWKGLFCGDG-GWLAATAKAINAIGTFLKSQ 177

QY 187 VKFVTGT-----PQSRVVTLIFENNVDKGTGDKIWAERTFLCAGASAGQL 236
Db 178 VKFGSGAGTKRPLFAPDGTCS-----GVETVDGTYKVPADKVVLAAGAWSSTLV 228

QY 237 DFKNQLRPTATWTLVHALKPEERALKYNIPIVFNIERGFFPEDEERGEIKICDEHPGYT 296
Db 229 DLEDQCVSKAWVAHQLTPQESAQYKDVVVDGYDGYGFFPEPN-EHGVTKVCDPEPGFS 287

QY 297 NM-----VQSGADGTMMSIPKTEQ-----IPKBAETRVALLKETMPQLADRPFSFARIC 346
Db 288 RFKLHPQYCATSPKLISVPSRSHAKPTDTPDSSEETIRKAIARFMPFKDKELFNRSKC 347

QY 347 WCADTANREFLIDRHPQYHSLVLCGASGRGFKYLPISIGNLIYDAMEGKVPQKIHILIKW 406
Db 348 WCTDTADANLLICEHPKWNKFIATGDSGHSFKVLPNIGKHVVVELIEGRLPQDLAGAWRW 407

QY 407 NP-----DIAANRNWRDTLGRFGGNRVMDFHVK 436
Db 408 RFGDALKSRAPAKDLAEMPGWK-----HDAK 436

RESULT 7
Q96UT4 PRELIMINARY; PRT; 438 AA.
AC Q96UT4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructosyl peptide oxidase (EC 1.5.3.-).

DE Fructosyl amino acid oxidase.
GN Name=faoA;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=FGSC A4;
RX MEDLINE=22261769; PubMed=12375102;
RA Jeong H.Y., Song M.H., Back J.H., Han D.M., Wu X., Monnier V.,
RA Jahng K.Y., Chae K.S.;
RT "The vea gene is necessary for the inducible expression by fructosyl
RT amines of the Aspergillus nidulans faoA gene encoding fructosyl amino
RT acid oxidase (amadoriase, EC 1.5.3).";
RL Arch. Microbiol. 178:344-350(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=FGSC A4;
RA Jeong H.-Y., Song M.H., Back J.H., Han D.M., Monnier V.M.,
RA Jahng K.-Y., Chae K.-S.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF416568; AAL09028.1; -.
DR InterPro; IPR006076; Fao_Oxred.
DR InterPro; IPR000205; NAD_BS.
DR Pfam; PF01266; DAO; 1.
SQ SEQUENCE 438 AA; 48433 MW; BF72BC2CC4D57316 CRC64;

Query Match 24.9%; Score 629; DB 2; Length 438;
Best Local Similarity 34.0%; Pred. No. 8.1e-40;
Matches 148; Conservative 73; Mismatches 168; Indels 46; Gaps 12;

QY 20 LLIVAGTGTGTSTLHLARRGT--NVTLDVPVPSPSAISAGNDVNVKVISSGQYSNNKDE 77
Db 10 IVGGGTGSGSTLHLARRGTTPSNITVLDTCPIPSAQSGYDLNKIMSI--RLNKPD 67

QY 78 IEVNEILAEAFNGWKNDPLFKPYVYHDTGLLMSACSQGLDRLGVRRP-----GEDPN 131
Db 68 LQ-----LSLEALDMWKNDPLFKPFHNVGMIDVSSTEEGIEGLRKKYQSILLDAGIGLEKT 123

QY 132 LVELTRPQFKLAPEGVLQDFFPGWKGYFARSGAGWAHARNALVAAAREAQRMGVKFTV 191
Db 124 NFMLESEDDILAKAPH-FTREQIKGWKGLFCGDG-GWLAATAKAINAIGTFLKEQGVKPGF 181

QY 192 GTPQGRVVTLIFENNVDK--GAVTGDGKIWAERTFLCAGASAGQLDKNQLRPTATW 248
Db 182 G-GAGTKFPLFADAHEKTCIGVEITVDGTYKVPADKVVLAAGAWSSTLVLEECVSKAWV 240

QY 249 LVHIALKPEERALKYNIPIVFNIERGFFPEDEERGEIKICDEHPGYTMVQ-----SAD 303
Db 241 FAHIQLTPAEEAAAYKNTPIYDGYDGYGFFPEPN-ENGIIKVCDFPGTFHFKMHQPYGSPA 299

QY 304 GTWMSIPPEKTIQPK-----AETRVALLKETMPQLADRPFSFARICWCADTANREFLI 358
Db 300 PKPISVPSRSHAKPTDTPHASEVTIKKAINRPLPRFNKDELNRAWCWCTDTADANLLV 359

QY 359 DRHPQYHSLVLCGASGRGFKYLPISIGNLIYDAMEGKVPQKIHILIKWPN-----408
Db 360 CEHPRWKGYLATGDSGHSFKVLPNIGKHVVVELLEERLESVFKDANWRPQSGDALKSR 419

QY 409 -----DIAANRNWRD 418
Db 420 AAPAKDLADMPGWRN 434

RESULT 8
Q765A9 PRELIMINARY; PRT; 437 AA.
AC Q765A9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Fructosyl peptide oxidase (EC 1.5.3.-).
```



```

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schafer M., Muller-Auer S.,
RA Gabel C., Fuchs M., Dusterhoft A., Fritz C., Holzer E., Moestl D.,
RA Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R.,
RA Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
DR EMBL; AL022071; CAA17815.1; -.
DR PIR; T40295; T40295.
DR GeneDB SPombe; SPBC354.15; -.
DR InterPro; IPR006076; Fad_oxred.
DR InterPro; IPR000205; NAD_BS.
DR Pfam; PF01266; DAO; 1.
DR SQUAD; 412 AA; 46319 MW; BD36B86075D3F4EF CRC64;
SQ
Query Match 19.3%; Score 488.5; DB 2; Length 412;
Best Local Similarity 30.1%; Pred. No. 4.8e-29;
Matches 130; Conservative 78; Mismatches 183; Indels 41; Gaps 12;
QY 14 VTKSSLLIVGAGTGTSTALHARGYNTVTVLDPPVPSAISAGNDVNVKISSGOYSN 73
DB 1 MVKNTSVIIVGAGVGLSAALETKGGYTIKILDRAPPVIDGSSVDANRIIRS-DYAD 59
QY 74 NKDEIEVNEILAEAFNGMKNDPLFKPYHYHDTGLMSA-----CSQEGDLRLGVVR 125
DB 60 -----AVYCSMGIDALEEWRNPLFKEQYFGSLMFVGRDNVEYRDMSLNLTGMGV--- 111
QY 126 PGEDPNLVELTRPEQFKLAPEGLVQDFGKYGKGFARSAGAHARNALVAARAEQRM 185
DB 112 -----SAAKFQTTTELRLKGLFKPMI--GELNDGAGYANFSSGWANAEQSVKSVVNYLAHA 164
QY 186 GVKFVTGTPOGRVTVLIFENNVDKGVATGDKIWRARTPLCAGASAGQFL-DFKNQLRP 244
DB 165 GVSVFISG-PGTYBELITEENVVKGVRTTIG-AYMAEKLIFFATGAWTASLLPNDHTRELA 222
QY 245 TAWTLVHIALKPEERALKYKNIPIVFNIERGFPEPDEERGEIKICDEHPGYTNVQSADG 304
DB 223 TQGPVAYIKLTPEEYIRFLTNPNVLDFTGTYIFPPPTPDGKLFARHGYGFTRMQNLKSG 282
QY 305 TWMISPEFTQI-----PKEATFVRALLKETM-PQLADRPFSFARICWCACDTANREFLID 359
DB 283 KVESVPKPKPLVSPILPKEAELDLRLNLTQTYGEEISQRFPYKTRICYDTTADAEFVFD 342
QY 360 RHPQYHSLVLCGASGRGFKYLPISGNLIVDAMEGKVPQKIHELKWK---NPDIARNNW 416
DB 343 YHPDYENLFTCTGSGHGFFPILGKYSIGCMFRELLEPLLKWRWKKENLEFALDHS 402
QY 417 RDTLGRFGGPNR 428
DB 403 R-----AGFSR 408
RESULT 11
Q9UTM9 PRELIMINARY; PRT; 433 AA.
AC Q9UTM9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE SPAC139.04c protein.

```

```

GN Name:SPAC139.04c;
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Spouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris A., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schafer M., Muller-Auer S.,
RA Gabel C., Fuchs M., Dusterhoft A., Fritz C., Holzer E., Moestl D.,
RA Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R.,
RA Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA McDougall R.C., Rajandream M.A., Barrell B.G., Saunders D., Harris D.,
RA Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL132667; CAB59618.1; -.
DR PIR; T37605; T37605.
DR GeneDB SPombe; SPAC139.04c; -.
DR InterPro; IPR006076; Fad_oxred.
DR InterPro; IPR000583; GATase_2.
DR Pfam; PF01266; DAO; 1.
DR PROSITE; PS00443; GATASE_TYPE_II; UNKNOWN 1.
DR SQUAD; 433 AA; 48973 MW; 170081BB3D48B1BE CRC64;
SQ
Query Match 18.4%; Score 465; DB 2; Length 433;
Best Local Similarity 30.1%; Pred. No. 3.3e-27;
Matches 130; Conservative 80; Mismatches 174; Indels 48; Gaps 13;
QY 17 SSSLIVGAGTGTSTALHAR-RGYNTVTVLDPPVPSAISAGNDVNVKISSGOYSNNK 75
DB 2 SRTIVIVGCVGLSTAVELAKNHSFDNIIAIDAEVPSPMSAANDINKIVRP-EYADLK 60
QY 76 DELEVNEILAEAFNGMKNDPLFKPYHYHDTGLMSACSQEGDLRLGVVRPGE----- 128
DB 61 -----YMKLALEAMEKWRNDPELSVYFEGRLST-----ISKPYRAFDEVAQRNL 109
QY 129 ----DPNLVELTRPEQFKLAP-----EGVLQGFPGWKGYFARSAGAHARNALVAAR 180
DB 110 KLGLDSALINLSSEIRKYPKPSLFSNSPLRSDMQA-----VNEHAGYANSAASKLLEL 165
QY 181 EAQRMGVKFTGTPQGR-----VV-----TLFENNVDKGVATGDKIWRARTPLCAGASA 232
DB 166 KARELGVEFVFG-KAGKFKFVNVHSETDIDKNDHVSQVQEDGTIYHADTILLAVGAYL 224
QY 233 GQFLDFKNQLRPTAWTLVHIALKPEERALKYKNIPIVFNIERGFPEPDEERGEIKICDEH 292
DB 225 NAYLNTSHRVCAKGLPVAHIQLTDEBFKTYKNMPIIFDPDCAAYAFPPYPTVKLIKLASTG 284

```

```

Qy 293 PGYTNMVSQ---ADGTMMSIPF---EKTQIPKEATRVALLKTMQOLADRPSPFARIC 346
Db 285 YEYVCNVEDYDENSQVSVIPHSQPSKSSLPKYAIQWRFLDTFLPLADRSLLNTKWC 344
Qy 347 WCADTANREFLIDRHPQVSHVLACGAGSGRGFKYLPISGNLIVDAMEGKVPQKHILIKW 406
Db 345 WISTEDANFLIDKVPQDFNVFVANGDSGHAFKFLPNIGRIYIAQRIILGDISSEWKDAWRW 404
Qy 407 NPDIANR-NWR 417
Db 405 REDDKASELKWR 416

RESULT 12
Q7S638
ID Q7S638 PRELIMINARY; PRT; 502 AA.
AC Q7S638;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 25-OCT-2004 (TrEMBLrel. 26, Last sequence update)
DE Hypothetical protein (Related to fructosyl amino acid oxidase).
GN Name=NCU04771.1; Synonyms=B15B10.150;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysseilis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Kryzstofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Varden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbold D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000331; EAA30984.1; -.
DR EMBL; BX897677; CAE85581.1; -.
DR InterPro; IPR006076; Fad_oxred.
DR Pfam; PF01266; DAO; 1.
KW Hypothetical protein.
SQ SEQUENCE 502 AA; 52930 MW; 39748F9E553D4FDE CRC64;

```

```

Query Match 16.98; Score 426.5; DB 2; Length 502;
Best Local Similarity 29.24; Pred. No. 3.7e-24;
Matches 138; Conservative 56; Mismatches 169; Indels 109; Gaps 21;

Qy 18 SSLIVGAGTGWGTSALHLARR-GYTNVTVL-----DPYVPSPSAISAGNDVNKVI 66
Db 30 SSLILGSGVGLTAYALSRHDFVNTTITVVDKPSNPDVFPAPDAASV--DTSRII 87

```

```

Qy 67 SSGOYNNKDEIEVNEILABEAFNGWK-----NDPLFKPYHYDTCGLMSACSQSG-LDRL 120
Db 88 RA-DYPD-----AAYAALAAEAQLQWRQTHDPDLGSEGRYSSEGLLVVA---DGPAPPP 138
Qy 121 GVRVRPG-----EDPNLV-----ELTRPEQFRKLAPGV 149
Db 139 GTPVAPGTSVIDKSLTGM DYARPSWANVLSSASSDPELAARIKECPNTDAIKALGTG- 197
Qy 150 LQGFPPGKGYFARSGAGWAAHARNALVAAREAQRMG-VKFPVTGTPOGRVVTLLIFENN DV 208
Db 198 --GSSGSW-GYI-NGLSGWANAGASMAWLKRVRAEGRINFVA---GEVTNLEVSGNTV 249
Qy 209 KGAVTGDKIWRARTFLCAGASAGQFLDPKQNLRPNTAWTLVHIALKPEERALKNIPIVI 268
Db 250 TGAFLSDGRVLSADLVNMSAGAWTGRVLDTGQATGQVLYIDLTPEEEAQLAHMEVI 309
Qy 269 FNTERGFPFDEBERGEIKICDEHPGYTNMVQSADGTWMSIP-----PEKT----- 314
Db 310 LNLSTGLFVIP-PRNGVLKVARHAYGYLN-----PTLSVPPPLASFTTTPAAAVSLPLT 362
Qy 315 -----QIPKEATRVALLKE--TMPQLADRPSPFARICWCADTANREFLIDRHPQYH 365
Db 363 TLTDPTLQIPTGEGADLLRRALHEMVPLPSLRDRPFKTRICWYSDTPTADPIVDYHPKYK 422
Qy 366 SLVLGCGASGRGPKYLPISGNLIVDAMEGKVPQKHILIKWNPDIANRNR 417
Db 423 GLFVATGDSGHAFKFLPVIKEKIADVIAGQCP-----PBFVGKWNMR 464

RESULT 13
Q6CLIO
ID Q6CLIO PRELIMINARY; PRT; 433 AA.
AC Q6CLIO;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to sp|Q9UTW9 Schizosaccharomyces pombe Putative fructosyl
DE amino acid oxidase.
GN ORFNames=KLIA0F02860g;
OS Kluyveromyces lactis NRRL Y-1140.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=284590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talia E.,
RA Goffard N., Frangeul L., Aigle M., Anthonard V., Babour A., Barbe V.,
RA Barney S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boistame A., Boyer J., Cattolico L., Confaniolieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikoleki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellens S., Potter S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382126; CAG97917.1; -.
DR InterPro; IPR006076; Fad_oxred.
DR InterPro; IPR000205; NAD_BS.
DR Pfam; PF01266; DAO; 1.
SQ SEQUENCE 433 AA; 48905 MW; 1091A3C6C3F3DA08 CRC64;

```


Search completed: May 27, 2005, 13:47:31
Job time : 104.659 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1995 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:28:05 ; Search time 179.139 Seconds
(without alignments)
1019.048 Million cell updates/sec

Title: US-10-622-893A-5

Perfect score: 2530

Sequence: 1 MGSQDDDLALAVTKSSL.....LEGIPNPPLRTGHHHHH 472

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1982.5	78.4	437	2	AAW39253 A. terreu
2	1982.5	78.4	437	2	AAI29100 A. terreu
3	1977.5	78.2	437	2	AAW37141 Aspergill
4	1158.5	45.8	440	2	AAW69251 Fructosyl
5	1158.5	45.8	440	5	ABG30782 Fusarium
6	1152.5	45.5	440	5	ABG30784 Fusarium
7	1151.5	45.5	440	5	ABG30785 Fusarium
8	1150.5	45.5	440	5	ABG30783 Fusarium
9	1115.5	44.1	444	8	ADM78789 Fusarium
10	837.5	33.1	427	8	ADS15771 Fructosyl
11	661	26.1	432	7	ADE81213 Orf20, SE
12	660	26.1	437	2	AAW24134 Fructosyl
13	660	26.1	437	2	AAW97386 A fructos
14	660	26.1	437	2	AAW97387 A fructos
15	627	24.8	437	6	ABR41990 Eupenicil
16	600	23.7	437	6	ABR41989 Coniochae
17	437.5	17.3	439	5	ABB78358 Amino aci
18	416.5	16.5	473	8	ADM78787 Fusarium
19	330.5	13.1	109	2	AAW69252 Fructosyl
20	198	7.8	411	7	ADC59522 Arthroba
21	192	7.6	390	1	AAW94663 New sarco
22	192	7.6	390	2	AAW07397 Bacillus
23	184	7.3	389	2	AAW23780 Sarcosine
24	181.5	7.2	440	4	ABB59342 Drosophil
25	177.5	7.0	390	4	ABB08949 Human ser

26	177.5	7.0	390	4	AAW40278 Human pol
27	177.5	7.0	390	4	AAW94212 Human pro
28	177.5	7.0	400	4	AAW42064 Human pol
29	171	6.8	389	2	AAW711462 A modifie
30	169	6.7	387	2	AAW38078 Sarcosine
31	169	6.7	389	2	AAW52658 Arthroba
32	169	6.7	389	2	AAW79150 Arthroba
33	169	6.7	389	2	AAW76735 N-methyl
34	169	6.7	389	2	AAW25148 Arthroba
35	169	6.7	389	2	AAW711461 A modifie
36	169	6.7	389	2	AAW53155 Arthroba
37	169	6.7	389	4	AAW99724 Arthroba
38	169	6.7	389	8	ADO23559 Arthroba
39	168	6.6	389	2	AAW25150 Modified
40	168	6.6	389	2	AAW25151 Modified
41	168	6.6	389	2	AAW53164 Arthroba
42	168	6.6	433	4	AAW98942 E. coli g
43	168	6.6	433	6	ABU14777 Protein e
44	167	6.5	389	2	AAW53163 Arthroba
45	165	6.5	385	6	ABU40301 Protein e

ALIGNMENTS

RESULT 1

AAW39253
ID AAW39253 standard; protein; 437 AA.

XX AC AAW39253;

XX AC AAW39253;

DT 14-MAY-1998 (first entry)

XX A. terreus FAOD-L protein.

DE A. terreus FAOD-L protein.

XX A. terreus FAOD-L protein.

KW Fructosylamino acid oxidase; FAOD-L; Amadori compound.

XX Fructosylamino acid oxidase; FAOD-L; Amadori compound.

OS Aspergillus terreus.

XX Aspergillus terreus.

PN JPI0033180-A.

XX JPI0033180-A.

PD 10-FEB-1998.

XX 10-FEB-1998.

PF 24-JUL-1996; 96JP-00194557.

XX 24-JUL-1996; 96JP-00194557.

PR 24-JUL-1996; 96JP-00194557.

XX 24-JUL-1996; 96JP-00194557.

PA (KYOT-) KYOTO DAIICHI KAGAKU KK.

XX (KYOT-) KYOTO DAIICHI KAGAKU KK.

DR WPI; 1998-172097/16.

DR N-PSDB; AAV09626.

XX N-PSDB; AAV09626.

PT Aspergillus terreus recombinant fructosyl:amino acid oxidase - used for

PT analysis of "Amadori" compound(s).

XX Example 1; Col 23-26; 18pp; Japanese.

PS Example 1; Col 23-26; 18pp; Japanese.

XX This sequence represents a novel recombinant fructosylamino acid oxidase

CC (FAOD-L) produced from eukaryotic cells integrated with an RT-PCR product

CC encoding FAOD-L derived from Aspergillus terreus GP1 (FERM P-15664). FAOD

CC -L can be used for the analysis of "Amadori" compounds

XX -L can be used for the analysis of "Amadori" compounds

SQ Sequence 437 AA;

Query Match 78.4%; Score 1982.5; DB 2; Length 437;

Best Local Similarity 82.2%; Pred No. 9.5e-188;

Matches 360; Conservative 44; Mismatches 33; Indels 1; Gaps 1;

Qy 12 LAVTKSSLLIVGAGTGTGTALHARRGTYNTVLDPPYVPSAISAGNDVKNVSSQY 71

Db 1 MPVTSSSLLIIGAGTGTGTALHARRGTYNTVLDPPYVPSAISAGNDVKNVSSQY 60

Qy 72 SNNKDETEVNEILLAEAFNGWKNDFPKPYHTDTGLLMSACSQEGDLRLGVRVPGEDPN 131

61	SSKODEVEYNEI	IABQAFNGWKNDIPFKPYVYDHTGVVMSATTQEGLERLGRVVRPEDEPD	120
Qy	132	LYVELTRPEQFRKLAPEGVLQDFFGCKGYPARSAGAGWAHARNALVAAAREAQRMGVKFTY	191
Db	121	VABLTRPEQFROLAR-GVYKGNFPQWRGYHIRSNAGWAHARNALVAAAREAQRLGRVFA	179
Qy	192	GTPQGRVVTLIPENNNDVGAVTGDGKIWRABRTFLCAGASAGOLFDPKQOLRPTAMTLVH	251
Db	180	GSFQGRVITLIPENNNDVGAVTADGKIWRABQTLTLCAGAAGOLFDPKQOLRPTAMTLVH	239
Qy	252	IALKPEERALKYNIPIVENIERGFFFPDEBERGEIKICDEHPGYTNVWSADATMMSIPF	311
Db	240	IQLKPEERAQYKMPVPVNIKGFPPFDEBERGEIKICDEHPGYTNNTGADGRVRSIPF	299
Qy	312	EKTQIPKEATRVRRALKETPQLADRPFSFARICWCADTANREFLIDRHPYHSLVLCG	371
Db	300	EKTQVPREAEVRKLLSETMPQLADRPFSFARICWCADTPNREFIIDRHPYPSLVLCG	359
Qy	372	GASGRGFKYLPISGNLIVDAMEGKVQPKIHელიKWNPDIAANRNWRDITLGRFGGPNRVM	431
Db	360	GASGRGFKYLPISGSIADAMEDKTPAKIHKLRWSPEIAINRNWGDRLGRFGGPNRVM	419
Qy	432	FHDVKEWTNVQYRDISKL	449
Db	420	FNEVKEWITNVQTDISKL	437

RESULT 2	
AAAY29100	
ID	AAAY29100 standard; protein; 437 AA.
XX	
AC	AAAY29100;
XX	
DT	02-NOV-1999 (first entry)
XX	
DE	A. terreus heat-resistant fructosyl amino acid oxidase (FAOD) .
XX	
KW	FAOD; thermostable; fructosyl amino acid oxidase; heat-resistant;
XX	glutathione S-transferase; GST; Amadori compound.
XX	
OS	Aspergillus terreus.
XX	
PN	JP11221081-A.
XX	
PD	17-AUG-1999.
XX	
PF	10-FEB-1998; 98JP-00028169.
XX	
PR	10-FEB-1998; 98JP-00028169.
XX	
PA	(KYOT-) KYOTO DAIICHI KAGAKU KK.
XX	
WPI	1999-520719/44.
DR	N-PSDB; AA207721.
XX	
PT	New heat-resistant fructosyl amino acid oxidase - useful for analysis of
XX	Amadori compounds.
PS	Example; Page 5-6; 8pp; Japanese.
XX	
CC	The present sequence represents an Aspergillus terreus heat-resistant
CC	(thermostable) fructosyl amino acid oxidase (FAOD) . FAOD can be produced
CC	using standard recombinant methodology. A recombinant fusion protein
CC	comprising FAOD and glutathione S-transferase (GST) is useful for the
CC	analysis of Amadori compounds
XX	
SQ	Sequence 437 AA;

Query Match	78.4%;	Score 1982.5;	DB 2;	Length 437;
Best Local Similarity	82.2%;	Pred. NO. 9.5e-188;		
Matches 360; Conservative	44;	Mismatches 33;	Indels 1;	Gaps 1;

Qy		12	LAVTKSSLLIVGAGTWTGTSTALHARRGYNTVTLDPYPVPISAI SAGNDVNKVISSGOY	71
Dd		1	MPVTKSSSIIIGAGTWGCSTALHARRGYNTVTLDPYPVPISAI SAGNDVNKI I SS GOY	60
Qy		72	SNNKDETEVNEIILAEEAFNGWKNDPLFKPYTHDTGLLMSACSOEGLDRLGVVRPCEDPN	131
Dd		61	SSSKDEVEVNEIIAQEPFGWKNDPIFKPYTHDTGVVNSATTOEGLERLGVVRPEDEPD	120
Qy		132	LVELTRPEQFRKLAPGLQDFPGWKGYFARSGAGWAHARNALVA AAREAQMGKFVFT	191
Dd		121	VAELTRPEQFROLAP - GVLKGNFPWGRGYHIRSNAGWAHARNALVA AAREAQRLGVRFVA	179
Qy		192	GTPQGRVVUTLI FENNDDKGAVTGDKIWRABRTFLCAGASAGOFLDFKNOLRPTAWTLVH	251
Dd		180	GSPQGRVITLIFENNDDKGVATADGKIWRABOQTILCAGAAGQFLDFKDQURPTAWTLVH	239
Qy		252	IALKPEBRALYNKIPVIFNIERGFFPFDEBERGEIKICDEHPGYTNMVQSADGTWMSIPF	311
Dd		240	IQLKPEERAQYKNMPVVFNIEKGFFFEDEBERGEIKICDEHPGYTNMTTGADGRVRSIPF	299
Qy		312	EKTQIPKEAETRVRA LLKETNPQLADRPFSFARI CWCADTANREPLIDRHPOYHSILVLCG	371
Dd		300	EKTQVPBEAEMRVRLKLS ETNPQLADRPFSFARI CWCADTNPREEI IDRHPEYPSLVLCG	359
Qy		372	GASGRGFYLPSIGNLVLDA MEGVKVPQKIH E L I K W N P D I A A N R N W D T L G R F G G P N R V M D	431
Dd		360	GASGRGFYLPSIGSI I ADAMEDKTPAKIHKLRWSPEIA INRWNGDRLGRFGGPNRVMD	419
Qy		432	FHDVKEWNTVOYRDISKL	449
Dd		420	FNEVKEWNTVORDISKL	437

RESULT 3	
AAW37141	
ID	AAW37141 standard; protein; 437 AA.
XX	
AC	AAW37141;
XX	
DT	17-OCT-2003 (revised)
DT	06-JUL-1998 (first entry)
XX	
DE	Aspergillus terreus fructosyl amino acid oxidase.
XX	
KW	Fructosyl amino acid oxidase; FAOD-L; amadori compound; assay;
KW	food analysis; diabetes; diagnosis.
XX	
OS	Aspergillus terreus; strain GP1 (FERM BP-5684).
XX	
PN	EP821064-A2.
XX	
PD	28-JAN-1998.
XX	
PF	19-JUL-1997; 97EP-00112403.
XX	
PR	23-JUL-1996; 96JP-00193344.
XX	
PA	(KYOT-) KYOTO DAIICHI KAGAKU CO LTD.
XX	
PI	Kato N, Sakai Y, Tani Y, Fukuya H;
XX	
DR	WPI; 1998-088887/09.
DR	N-PSDB; AAV00732.
XX	
PT	Recombinant fructosyl amino acid oxidase - useful in assays for amadori
PT	compounds, e.g. in diabetes diagnosis or food analysis.
XX	
PS	Claim 1; Page 20-21; 33pp; English.
XX	
CC	This polypeptide comprises an Aspergillus terreus fructosyl amino acid
CC	oxidase (FAOD), designated FAOD-L, that shows higher FAOD activity on
CC	fructosyl lysine than on fructosyl valine. Its amino acid sequence was
CC	deduced from a cDNA clone (see AAV00732) obtained from an A. terreus GPI
CC	

CC in on glycoamino acids. This sequence represents a polypeptide used
 CC in the scope of the invention
 XX
 SQ Sequence 440 AA;

Query Match 45.6%; Score 1152.5; DB 5; Length 440;
 Best Local Similarity 50.9%; Pred. No. 3.6e-105;
 Matches 223; Conservative 75; Mismatches 133; Indels 7; Gaps 5;

Qy 11 ALAVTKSSLLIVGAGTGTSTALHLARRGYTNVTLDPYPVPSAISAGNDVNVKVIS--S 68
 Db 1 ASTLTQSQILIVGGGTWGCSTALHLARRGYTNVTLDPVPSAISAGHDVNVKLAGRLS 60
 Qy 69 GOYSNNKDEIEVNEILAEAFNGWKNPDLKPYHYDTGLLMSACSQGLDLRGLVVRPGE 128
 Db 61 TADSKGDEDSIWKALSAAAQGLHDPVFPFCHNTGTSVVGSTPKSIKQL-VEDEIGD 119
 Qy 129 D-PNLVELTRPEQFRKLAPGVLQDPPGKMGYFARSAGAWAHARNALVAAAREARMGV 187
 Db 120 DIDQYPLNTAEDFRKTMPEGILTGNPFGWKGYKPTGSGVHARKAKAAFESEERLGV 179
 Qy 188 KFTVTGPGQRVWTLIFENNNDVKGAVTDGDKIWAERTFLCAGASAGQFLDKQLRPTAW 247
 Db 180 KFTTSGPEGKVESLIFEDGDVGRGAKTADGKEHRADRTILSAGASAEFFLDFENQIQPTAW 239
 Qy 248 TLVHIALKPEERALKYKNIPIVNIERGEFFPEDEERGEIKICDEHPGYTNMVQADGTM- 306
 Db 300 QSIIPFAKHQVPTAERRMKQFLKIDIMPQLADRLPLVHARICWCADTQDRMFLITYHPHPS 359
 Qy 367 LVLGCGASGRGFKYLSIGNLIVDAMEGKVPQKIHILIKWNPDIANRNRDITLGRFGGP 426
 Db 360 LVIASGDCGTYVHITSIGKIFSDCMEGTLEERFAKYWRWRPEKFTFEGWKDPLDRFGAD 419
 Qy 427 NRWVDF--HDVKWNTNVQ 442
 Db 420 DKIMDLPKSDVEGWNTIK 437

RESULT 7
 ABG30785
 ID ABG30785 standard; protein; 440 AA.
 AC ABG30785;
 XX
 DT 05-NOV-2002 (first entry)
 XX
 DE Fusarium oxysporum polypeptide #4.
 XX
 KW Glycoamino acid; glycoprotein; deoxycholic acid; deoxycholic amide;
 KW cholic acid amide octyl glucoside; quaternary ammonium salt;
 KW quaternary ammonium salt cationic surfactant; concanavalin A; betaine;
 KW ascorbic oxidase; 4-(2-hydroxyethyl)-1-piperazinyl group; albumin;
 KW protease; glycoalbumin; bromocresol purple; glycation; diabetes;
 KW ascorbic acid; mutant; mutein.
 XX
 OS Fusarium oxysporum.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 246
 FT /note= "Encoded by ATG"
 FT Misc-difference 372
 FT /note= "Wild-type Lys substituted by Val"
 FT Misc-difference 396
 FT /note= "Encoded by TTC"
 XX
 PN WO200261119-A1.
 XX
 PD 08-AUG-2002.

XX
 FF
 PR 31-JAN-2001; 2002WO-JP000721.
 PR 16-FEB-2001; 2001JP-00022953.
 PR 08-AUG-2001; 2001JP-00039796.
 XX 08-AUG-2001; 2001JP-00240002.
 PA (ASAH) ASAH KASEI KOGYO KK.
 XX
 PI Kouzuma T, Yoshioka I, Arai M, Sumitani J, Imamura S;
 XX
 XX WPI; 2002-599854/64.
 DR N-PSDB; ABK30478.
 XX
 PT Compositions for assaying glycoprotein without interference from globulin
 PT and ascorbic acid in blood, useful in clinical examination including
 PT diagnosis and management of diabetes.
 XX
 PS Disclosure; Page 72-75; 82pp; Japanese.
 XX
 CC The invention relates to compositions for assaying glycoprotein with use
 CC of protease and enzymes acting on glycoamino acids in the presence of at
 CC least 1 of deoxycholic acid or amide, cholic acid amide octyl glucoside,
 CC quaternary ammonium salt, quaternary ammonium salt cationic surfactant, a
 CC concanavalin A and betaine and/or ascorbic oxidase and a buffer without a
 CC 4-(2-hydroxyethyl)-1-piperazinyl group. The invention also relates to
 CC compositions for assaying albumin containing the protease and enzymes for
 CC determining glycoalbumin and a protein-denaturing agent and/or a compound
 CC having S-S bond and bromocresol purple for evaluation of glycation
 CC proportion of albumin into glycoalbumin with respect to the other
 CC albumins measured separately. The compositions are useful in clinical
 CC examination including diagnosis and management of diabetes. The
 CC compositions can be used in methods which avoid the effects of globulin
 CC and ascorbic acid components and stabilise enzymes such as protease
 CC acting on glycoamino acids. This sequence represents a polypeptide used
 CC in the scope of the invention
 XX
 SQ Sequence 440 AA;

Query Match 45.5%; Score 1151.5; DB 5; Length 440;
 Best Local Similarity 50.9%; Pred. No. 4.5e-105;
 Matches 223; Conservative 75; Mismatches 133; Indels 7; Gaps 5;

Qy 11 ALAVTKSSLLIVGAGTGTSTALHLARRGYTNVTLDPYPVPSAISAGNDVNVKVIS--S 68
 Db 1 ASTLTQSQILIVGGGTWGCSTALHLARRGYTNVTLDPVPSAISAGHDVNVKLAGRLS 60
 Qy 69 GOYSNNKDEIEVNEILAEAFNGWKNPDLKPYHYDTGLLMSACSQGLDLRGLVVRPGE 128
 Db 61 TADSKGDEDSIWKALSAAAQGLHDPVFPFCHNTGTSVVGSTPKSIKQL-VEDEIGD 119
 Qy 129 D-PNLVELTRPEQFRKLAPGVLQDPPGKMGYFARSAGAWAHARNALVAAAREARMGV 187
 Db 120 DIDQYPLNTAEDFRKTMPEGILTGNPFGWKGYKPTGSGVHARKAKAAFESEERLGV 179
 Qy 188 KFTVTGPGQRVWTLIFENNNDVKGAVTDGDKIWAERTFLCAGASAGQFLDKQLRPTAW 247
 Db 180 KFTTSGPEGKVESLIFEDGDVGRGAKTADGKEHRADRTILSAGASAEFFLDFENQIQPTAW 239
 Qy 248 TLVHIALKPEERALKYKNIPIVNIERGEFFPEDEERGEIKICDEHPGYTNMVQADGTM- 306
 Db 300 QSIIPFAKHQVPTAERRMKQFLKIDIMPQLADRLPLVHARICWCADTQDRMFLITYHPHPS 359
 Qy 367 LVLGCGASGRGFKYLSIGNLIVDAMEGKVPQKIHILIKWNPDIANRNRDITLGRFGGP 426
 Db 360 LVIASGDCGTYVHITSIGKIFSDCMEGTLEERFAKYWRWRPEKFTFEGWKDPLDRFGAD 419
 Qy 427 NRWVDF--HDVKWNTNVQ 442
 Db 420 DKIMDLPKSDVEGWNTIK 437

Db 420 DKIMDLPKSDVEGWNIK 437

RESULT 8

ABG30783

ID ABG30783 standard; protein; 440 AA.

XX

AC ABG30783;

DT 05-NOV-2002 (first entry)

XX

DE Fusarium oxysporum polypeptide #2.

XX

KW Glycino acid; glycoprotein; deoxycholic acid; deoxycholic amide;

KW cholic acid amide octyl glucoside; quaternary ammonium salt;

KW quaternary ammonium salt cationic surfactant; concanavalin A; betaine;

KW ascorbic oxidase; 4-(2-hydroxyethyl)-1-piperaziny group; albumin;

KW protease; glycoalbumin; bromocresol purple; glycation; diabetes;

XX ascorbic acid; mutant; muten.

OS Fusarium oxysporum.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 246 /note= "Encoded by ATG"

FT Misc-difference 372 /note= "Wild-type Lys substituted by Trp. Encoded by TTC"

FT Misc-difference 396 /note= "Encoded by TTC"

FT

XX W0200261119-A1.

XX

PD 08-AUG-2002.

XX

PF 30-JAN-2002; 2002WO-JP000721.

XX

PR 31-JAN-2001; 2001JP-00022953.

PR 16-FEB-2001; 2001JP-00039796.

PR 08-AUG-2001; 2001JP-00240002.

XX

PA (ASAH) ASAH KASEI KOGYO KK.

XX

PI Kouzuma T, Yoshioka I, Arai M, Sumitani J, Imamura S;

XX

DR WPI; 2002-599854/64.

DR N-PSDB; ABK90476.

XX

XX Compositions for assaying glycoprotein without interference from globulin and ascorbic acid in blood, useful in clinical examination including diagnosis and management of diabetes.

PS Disclosure; Page 66-69; 82pp; Japanese.

XX

CC The invention relates to compositions for assaying glycoprotein with use of protease and enzymes acting on glycoamino acids in the presence of at least 1 of deoxycholic acid or amide, cholic acid amide octyl glucoside, quaternary ammonium salt, quaternary ammonium salt cationic surfactant, concanavalin A and betaine and/or ascorbic oxidase and a buffer without a 4-(2-hydroxyethyl)-1-piperaziny group. The invention also relates to compositions for assaying albumin containing the protease and enzymes for determining glycoalbumin and a protein-denaturing agent and/or a compound having S-S bond and bromocresol purple for evaluation of glycation proportion of albumin into glycoalbumin with respect to the other albumins measured separately. The compositions are useful in clinical examination including diagnosis and management of diabetes. The compositions can be used in methods which avoid the effects of globulin and ascorbic acid components and stabilise enzymes such as protease acting on glycoamino acids. This sequence represents a polypeptide used in the scope of the invention

XX

SQ Sequence 440 AA;

Query Match 45.5%; Score 1150.5; DB 5; Length 440;
Best Local Similarity 50.9%; Pred. No. 5.7e-105;
Matches 223; Conservative 75; Mismatches 133; Indels 7; Gaps 5;

QY 11 ALAVTKSSLLIVGAGTGTSTALHARRGYTNVTVDLPYPVPSAISAGDNVKNKVL--S 68
DB 1 ASTLTQSQILIVGGTGWGCTALHARRGYTNVTVDLPYPVPSAISAGDNVKNKLAGRLS 60
QY 69 GOYSNNKDEIEVNEJLAEEAFNGKNDPLFKPYKHTDGLLMSACSOEGDLRLGVRVPGE 128
DB 61 TADSKGDDSDSIWALSAYAAQGLWLDVPFQPCHTGVSAGSTPKSIKQL-VEDEIGD 119
QY 129 D-PNLVELTRPEQFRKLAPEGLVQGDFFGKMGYFARSAGWAGHARNALVAAREAGRMGV 187
DB 120 DIDQVTPLNTAEDFRKTMPEGILTGNFGKMGYFARSAGWAGHARNALVAAREAGRMGV 179
QY 188 KFTVTGPQGRVTLIFENNNDVKGAVTGDKIWRABRTFLCAGASAGQFDFKNQLRPTAW 247
DB 180 KFTTSGPEGKVESLIFEDGDVGRGAKTADGKRAHRTILSAGASAEFFLDFENQIOPTAW 239
QY 248 TLVHIALKPEERALYKNIPIVFNIERGFFEPDEERGEIKICDEHPGYTMVQSADGTM- 306
DB 240 TLGHQITPEETKLYKNLPFLFNINQGFWEDEDLHQLKMCDEHPGCYCNWEKPSKYP 299
QY 307 MSIPPEKTOIPKEAETRVALLKETMPQLADRPSPARICWCADTANRBFIDRHPOYHS 366
DB 300 QSIPFAKHQVPTAEARRMKQFLKIMPQLADRPVHARICWCADTQDRMFLITYHPRHS 359
QY 367 LVLGCGSGRGFKYLPISIGNLIVDAMEGKVPQKIHELKWNPDIAANRWDTLGRFGGP 426
DB 360 LVIASGDCGTGYWHITSIGKFSIDCMEGTLEERFAKYWRWRPEKFTFEGKODPLDRFGAD 419
QY 427 NRVMDF--HDVKWNTVQ 442
DB 420 DKIMDLPKSDVEGWNIK 437

RESULT 9

ADM78789

ID ADM78789 standard; protein; 444 AA.

XX

AC ADM78789;

XX

DT 01-JUL-2004 (first entry)

XX

DE Fusarium proliferatum fructosylamine oxidase Q2 protein sequence.

XX

KW fructosylamine oxidase; fructosyllysine; fructosylvaline;

KW Amadori compound; medical examination; diagnosis; serum glucose level;

KW food inspection; enzyme.

XX

OS Fusarium proliferatum.

XX

PN W02004029251-A1.

XX

PD 08-APR-2004.

XX

PF 16-SEP-2003; 2003WO-JP011766.

XX

PR 24-SEP-2002; 2002JP-00277214.

PR 24-OCT-2002; 2002JP-00309734.

XX

PA (ARKR-) ARKRAY INC.

XX

PI Yoshida N, Tani Y, Yonehara S;

XX

DR WPI; 2004-316127/29.

DR N-PSDB; ADM78788.

XX

PT Fusarium proliferatum-originated fructosylamine oxidase useful in measurement of Amadori compound and particularly applicable in medical examination including diagnosis and management of diabetes.

XX

Db	288	RFKLHQPVCATSPKLI	SVPRSHAKUPTDTY	PDSSSEETIRKATARPMPRPFKDKELFNRS	MC 344		
Qy	347	WCADTANREFILDRHPQYHSLV	GLGCAGSRGFKYLP	PSIGNLIVDAWEGKVPQKIH	ELIKW 406		
Db	348	WCTDTADANLLICEHPKWKNFILAT	GDGSHSPKVL	PNIGKHVVELIEGRLPQDL	AGAWRW 407		
Qy	407	NP-----DIAANRNWRD	TILGRFCGPNRVMD	FHDVK	436		
Db	408	RPGGDALKSR	SAPAKDLAEMP	GWK-----HDAK	436		
RESULT 13							
AAW97386							
ID	AAW97386	standard; protein; 437 AA.					
XX	XX	AAW97386;					
XX	XX	14-MAY-1999 (first entry)					
XX	XX	A fructosylamino acid oxidase.					
XX	XX	Fructosylamino acid oxidase; FAOD-P; Penicillium janthinellum S-3413.					
XX	XX	Synthetic.					
OS	OS	Penicillium janthinellum.					
XX	XX	JP11046769-A.					
XX	XX	23-FEB-1999.					
XX	XX	05-AUG-1997; 97JP-00210609.					
XX	XX	05-AUG-1997; 97JP-00210609.					
XX	XX	(KYOT-) KYOTO DAIICHI KAGAKU KK.					
XX	XX	WPI; 1999-208112/18.					
DR	DR	N-PSDB; AAX15949.					
XX	XX	New DNA coding fructosylamino acid oxidase - synthetically designed and recombinantly prepared.					
PT	PT	Claim 1; Page 8-10; 2lpp; Japanese.					
XX	XX	The present sequence represents a fructosylamino acid oxidase (FAOD-P), and is encoded by a synthetic sequence. The wild type FAOD-P sequence is derived from Penicillium janthinellum S-3413. The synthetic sequence was constructed using primers AAX15951-6006					
XX	XX	Sequence 437 AA;					
Query Match 26.1%; Score 660; DB 2; Length 437;							
Best Local Similarity 33.6%; Pred. No. 3.5e-56;							
Matches 156; Conservative 75; Mismatches 157; Indels 76; Gaps 14							
Qy	16	KSSLLIV-GAGTGTCTALHLAR	GYT--NVT	LDPPVPVPSAISAGNDV	NKVISGGVS 72		
Db	6	ESTKIVVGSGGTWGSSTALHL	IRSGYTPS	NTVLDVYPIPSLQ	SAGYDLNKMIS--RL 63		
Qy	73	NKDEIEVNEILAEAFNGWKN	NDPLFKPYHDTGL	LLMSACSGEGLDR	LGVRVRPGE	DNL 132	
Db	64	RNGPDLQ---	LSLEALDMWKNDPLFKP	FFHNVMGLDCSS	OEGTASLRKHQDL	LIDANI 119	
Qy	133	-----VELTRPEQPRKLA	PEGVLQDPFGWKG	YFARSGAGWAHARNAL	VAAAREAQRM	186	
Db	120	GLEKTNWLESEDDILAKAPH-	FTREQIKGWKGLFCGDG-	GWLA	AAKAINAIGTFLK	SGQ 177	
Qy	187	VKFVTGT-----PQGR	VVTILIFENNDVK	GAVTGDKI	WRAERTFLCAG	SAGOP 236	
Db	178	VKFGGSGAGTFRPLFAPDG	TCS-----GVET	VDGTYK	FADKVVLAAG	WSSTLV 228	
Qy	237	DFKNQLRP	TAWTLVHIA	LKPEER	ALYKNIPI	VNIETRGFFFPDPD	ERGEIKICDHPGYT 296

QY	237	DFKNQLRPTAWTLVHIAIKPBERALYNKVPVFNIERGFFPEDEDERGEIKICDEHPGVT	296
Dd	229	DLEDDQCCKWAWFAHIQLTPOESAQYKDVPVVYDGYGFFFEPN-EHGVIKVCDEFPGFS	287
QY	297	NM-----VQSADGTWMSIPFEKTO-----IPKEAETRVALLKETMPQLADRPFSPARIC	346
Dd	288	RFLKHQPYGATSPKLISVPRSHAKHTDTYDPSSEETIRKAIRAFMRFKDKELFNRSMC	347
QY	347	WCADTANREFLIDRHQPVHSILVLCGASGRGFKYLPISGLNIVDAMEGKVYPQKIHELIKW	406
Dd	348	WCTDITADANLLICEHPKKNFILATGDSCHSFVKLPINIGHVVELTEGRLPQDLAGAWRW	407
QY	407	NP-----DIAANNWRDITLGREGGNRMVDHDYK	436
Dd	408	RPGDALSKRSAPAKDLAEMPGWK-----HDAC	436
 RESULT 15 ABR41990			
ID	ABR41990	standard; protein; 437 AA.	
XX	ABR41990;		
XX	AC		
XX	DT		
XX	11-AUG-2003	(first entry)	
DE	Eupenicillium terrenum fructosyl peptide oxidase.		
DE	Fructosyl peptide oxidase; enzyme; diabetes; diagnosis.		
KW			
XX	Eupenicillium terrenum.		
OS			
PX	EP1291416-Al.		
NN			
XX	12-MAR-2003.		
PD			
XX			
XX	03-SEP-2002; 2002EP-00256108.		
PF			
XX	04-SEP-2001; 2001JP-00266665.		
PR	12-DEC-2001; 2001JP-00378151.		
PR	06-AUG-2002; 2002JP-00228727.		
XX			
XX	(KIKK) KIKKOMAN CORP.		
PA			
XX			
PI	Kurosawa K, Hirokawa K, Kajiyama N;		
XX			
DR	WPI; 2003-395517/38.		
XX	N-PSDB; ACC48873.		
DR			
XX	New fructosyl peptide oxidases useful as enzymes for clinical diagnosis.		
PT			
XX			
PS	Claim 20; Page 31-34; 59pp; English.		
XX			
CC	The present sequence is the protein sequence of a novel fructosyl peptide		
CC	oxidase (FPO) of Eupenicillium terrenum ATCC18547. FPO enzymes of the		
CC	invention act on fructosyl valyl histidine in the presence of oxygen and		
CC	catalyse a reaction that produces alpha-ketoaldehyde, valyl histidine and		
CC	hydrogen peroxide. The E. terrenum FPO exhibits 182% relative activity		
CC	for fructosyl glycine and 9.78% for epsilon fructosyl lysine when		
CC	compared to 100% activity for fructosyl valyl histidine. It shows optimal		
CC	activity in the pH range 6.0-8.0 (preferably pH 7.0) and at temperatures		
CC	around 35 degrees C, and is stable in the pH range 5.0-9.0 and at		
CC	temperatures up to about 45 degrees C. The Km value for fructosyl valyl		
CC	histidine is 4.25 mM. The E. terrenum cDNA can be used for mass		
CC	production of the enzyme by recombinant methods. Stable FPO enzymes such		
CC	as E. terrenum FPO are useful in clinical diagnosis, e.g. of diabetes		
XX			
SQ	Sequence 437 AA:		

Qy	12	LAVTKSSLLIV--GAGTWTGSTALHARGYT--NVTVLDPYPVPSAISAGNDVNKVIS	67
Db	1	MAHSRASTKVVVGGGGETIGSSSTALHLIRSYTSPNITVLDDVYKTPSLOSAGHDJLNKIM-	59
Qy	68	SGOYSNNKDEIVNEIILAEAFNGWKNDLPFKPYHYDHTGLLMSACSQBGLDRLGVRVRP-	126
Db	60	-GITLRNGPDLO---LSLESLDWMQNDLFPKPFHQVGMIDCSSKSGIENLRKYQTL	114
Qy	127	-----GEDPNLVELTRPEQFRKLAPEGVLOQDFPGWKGYFARSAGAGWAHARNALVAARE	181
Db	115	LDAGIGLEKTNVWVLESEDEILAKAP-NFTREQVKGWGLFCTDQ-GWLAATAKAINAIGTF	172
Qy	182	AORMGVKVFCTPGOVRVTLIF--ENNDVKGAVTGDGKIWEABRTFLCAGASAGQFLDPK	239
Db	173	LQDKGVFGFG-GAGTQQPFAADGKTCTIGLETTDGTKYFADKVVLAAGWASPTLVDDLE	231
Qy	240	NQLRPTAWTLVHIALKPEERALYKNIPIVFNIERGFPEPDEERGEIKICDEHPGYTNM-	298
Db	232	DQCVSKAWVFAHIQLTPKEADAYKNVPVYDGYGFFPEPN-EYGVIKVCDPEFGFSRPK	290
Qy	299	-----VQSGADGTMSIPEKTOIPKE-----AETVRALLKETMPQADRPSPFARICWCA	349
Db	291	LHOPYGAASPKNISVPSHAKHPDTTVPDASEVTIRKAIARFLPEPFKDELFRNTRMCWT	350
Qy	350	DTANREFILDRHPYHSLVLCGGASGRGFKYLPISIGNLIVDAMEGKVPQKHILKKNPD	409
Db	351	DTADANLLICEHPKKNKFIATGDSGHSFKLLPNLIGKHVELLEGSLQEMAGAWRRPG	410
Qy	410	IAANRNWRDTLGRFGGPNRVMDFHVKEW	438
Db	411	GDALRS-----RRGAPAK--DLAEMPGW	431

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:35:35 ; Search time 36.2303 Seconds
(without alignments)
972.511 Million cell updates/sec

Title: US-10-622-893A-5

Perfect score: 2530

Sequence: 1 MGSGDDDDALAVTKSSSL.....LEGPIPNLRTGHHHHH 472

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1982.5	78.4	437	2	US-09-031-059-1
2	1982.5	78.4	437	2	US-09-031-059-3
3	168	6.6	433	4	US-09-711-164-412
4	155	6.1	387	3	US-09-457-302-1
5	150	5.9	866	4	US-09-949-016-10219
6	146.5	5.8	448	4	US-09-252-991A-21043
7	136.5	5.4	383	4	US-09-893-096-5
8	135.5	5.4	378	3	US-09-199-637A-367
9	135.5	5.4	419	4	US-09-252-991A-21423
10	134.5	5.3	484	4	US-09-252-991A-24053
11	125	4.9	652	4	US-09-252-991A-30872
12	124.5	4.9	476	4	US-09-489-039A-12890
13	123.5	4.9	486	4	US-09-684-405-2
14	121.5	4.8	669	4	US-09-213-888-25
15	121.5	4.8	669	4	US-09-328-877D-25
16	117.5	4.6	374	4	US-09-468-647A-118
17	116.5	4.6	431	1	US-08-391-339-18
18	116.5	4.6	431	1	US-08-484-274A-18
19	115.5	4.6	568	4	US-09-266-965-135
20	115	4.5	505	4	US-09-902-540-9986
21	114.5	4.5	474	4	US-09-543-681A-7499
22	113.5	4.5	635	4	US-09-543-681A-7499
23	113.5	4.4	477	4	US-09-252-991A-27656
24	112.5	4.4	626	4	US-09-331-568A-27
25	110.5	4.4	375	4	US-09-328-352-6891
26	108	4.3	380	4	US-09-489-039A-7473
27	108	4.3	430	1	US-08-391-339-5

Sequence 5, Appli
Sequence 9, Appli
Sequence 7036, Ap
Sequence 81, Appl
Sequence 81, Appl
Sequence 14, Appl
Sequence 8, Appli
Sequence 2, Appli
Sequence 7529, Ap
Sequence 5943, Ap
Sequence 8, Appli
Sequence 18996, A
Sequence 32, Appl
Sequence 1, Appli
Sequence 26523, A
Sequence 12, Appl
Sequence 80, Appl
Sequence 20386, A

ALIGNMENTS

RESULT 1

US-09-031-059-1
; Sequence 1, Application US/09031059
; Patent No. 5948659
; GENERAL INFORMATION:
; APPLICANT: KATO, NOBUO
; APPLICANT: SAKAI, YASUYOSHI
; APPLICANT: TANI, YOSHIKI
; APPLICANT: FUKUYA, HIROSHI
; TITLE OF INVENTION: RECOMBINANT FRUCTOSYL AMINO ACID
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,059
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/899,172
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, MARC S
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 0020-4253P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-031-059-1

Query Match 78.4%; Score 1982.5; DB 2; Length 437;
Best Local Similarity 82.2%; Pred. No. 3.4e-194;
Matches 360; Conservative 44; Mismatches 33; Indels 1; Gaps 1;

```
QY 12 LAVTKSSLLIVGAGTWTGTSTALHLARRGYTNVTVDLPYPVPSAISAGNDVNKVISSGOY 71
DB 1 MPVTKSSLLIIGAGTWTGCGSTALHLARRGYTNVTVDLPYPVPSAISAGNDVNKIISSGOY 60
QY 72 SNNKDEIEVNEILAEAFNGWKNDPLFKPYHYDHTGGLMSACSGEGLDRLGVRVPGEDPN 131
DB 61 SSKKDEVEVNEIIAEQAFNGWKNDPLFKPYHYDHTGVVMSATTQEGLERLGVRVRPEDEPD 120
QY 132 LVELTRPEQKRLAPGVLQGGPPGKGYFARSGAGWAHARNALVAAAREAQRMGKVFY 191
DB 121 VAELTRPEQROLAP-GVLKGNFPGWRGYHRSNAGWAHARNALVAAAREAQRLGVRVFA 179
QY 192 GTPQGRVVTLLIFENNVDKAVTGDGKIWRAERTFLCAGASAGQFLDFKNQLRPTAWTLVH 251
DB 180 GSPQGRVITLLIFENNVDKAVTADGKIWRAEQTLICAGAAAGQFLDFKQQLRPTAWTLVH 239
QY 252 IALKPEERALKYKNIPIVFNIERGFFPEDEERGEIKICDEHPGYTNMVSADGTMMISIPF 311
DB 240 IQLKPEERAOYKMPVVFNIIEKGFFPEDEERGEIKICDEHPGYTNMTTGADGRVRSIPF 299
QY 312 EKTQIPKEAETRVALLKETMPQLADRPFSFARICWCADTANREFLIIDRHPOVHSLVLCG 371
DB 300 EKTQVPREAEMVRKLLSETMPQLADRPFSFARICWCADTANREFLIIDRHPEYPSLVLCG 359
QY 372 GASGRGFKYLPSTGNLIVDAMEGKVPQKIHELKWNPDIAANNRWDTLGRFGGPNRVM 431
DB 360 GASGRGFKYLPSTGSIADAMEDKTPAKIHKLRWSPEIARNWGDRLGRFGGPNRVM 419
QY 432 FHDVKWNTVQYRDISK 449
DB 420 FNEVKWNTVQYRDISK 437
```

RESULT 2

US-09-031-059-3

; Sequence 3, Application US/09031059

; Patent No. 5948659

; GENERAL INFORMATION:

; APPLICANT: KATO, NOBUO

; APPLICANT: SAKAI, YASUYOSHI

; APPLICANT: TANI, YOSHIKI

; APPLICANT: FUKUYA, HIROSHI

; TITLE OF INVENTION: RECOMBINANT FRUCTOSYL AMINO ACID

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH

; STREET: PO BOX 747

; CITY: FALLS CHURCH

; STATE: VA

; COUNTRY: USA

; ZIP: 22040-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/031,059

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/899,172

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: WEINER, MARC S

; REGISTRATION NUMBER: 32,181

; REFERENCE/DOCKET NUMBER: 0020-4253P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 205-8000

; TELEFAX: (703) 205-8050

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 437 amino acids

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-031-059-3
```

Query Match 78.4%; Score 1982.5; DB 2; Length 437;

Best Local Similarity 82.2%; Pred. No. 3.4e-194; Indels 1; Gaps 1;
Matches 360; Conservative 44; Mismatches 33;

QY 12 LAVTKSSLLIVGAGTWTGTSTALHLARRGYTNVTVDLPYPVPSAISAGNDVNKVISSGOY 71

DB 1 MPVTKSSLLIIGAGTWTGCGSTALHLARRGYTNVTVDLPYPVPSAISAGNDVNKIISSGOY 60

QY 72 SNNKDEIEVNEILAEAFNGWKNDPLFKPYHYDHTGGLMSACSGEGLDRLGVRVPGEDPN 131

DB 61 SSKKDEVEVNEIIAEQAFNGWKNDPLFKPYHYDHTGVVMSATTQEGLERLGVRVRPEDEPD 120

QY 132 LVELTRPEQKRLAPGVLQGGPPGKGYFARSGAGWAHARNALVAAAREAQRMGKVFY 191

DB 121 VAELTRPEQROLAP-GVLKGNFPGWRGYHRSNAGWAHARNALVAAAREAQRLGVRVFA 179

QY 192 GTPQGRVVTLLIFENNVDKAVTGDGKIWRAERTFLCAGASAGQFLDFKNQLRPTAWTLVH 251

DB 180 GSPQGRVITLLIFENNVDKAVTADGKIWRAEQTLICAGAAAGQFLDFKQQLRPTAWTLVH 239

QY 252 IALKPEERALKYKNIPIVFNIERGFFPEDEERGEIKICDEHPGYTNMVSADGTMMISIPF 311

DB 240 IQLKPEERAOYKMPVVFNIIEKGFFPEDEERGEIKICDEHPGYTNMTTGADGRVRSIPF 299

QY 312 EKTQIPKEAETRVALLKETMPQLADRPFSFARICWCADTANREFLIIDRHPOVHSLVLCG 371

DB 300 EKTQVPREAEMVRKLLSETMPQLADRPFSFARICWCADTANREFLIIDRHPEYPSLVLCG 359

QY 372 GASGRGFKYLPSTGNLIVDAMEGKVPQKIHELKWNPDIAANNRWDTLGRFGGPNRVM 431

DB 360 GASGRGFKYLPSTGSIADAMEDKTPAKIHKLRWSPEIARNWGDRLGRFGGPNRVM 419

QY 432 FHDVKWNTVQYRDISK 449

DB 420 FNEVKWNTVQYRDISK 437

RESULT 3

US-09-711-164-412

; Sequence 412, Application US/09711164

; Patent No. 6589738

; GENERAL INFORMATION:

; APPLICANT: Forsyth, R. Allyn

; APPLICANT: Ohlsen, Karl

; APPLICANT: Zyskind, Judith

; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THEREOF

; FILE REFERENCE: ELITRA.008A

; CURRENT APPLICATION NUMBER: US/09/711,164

; CURRENT FILING DATE: 2000-11-09

; PRIOR APPLICATION NUMBER: US 60/164415

; PRIOR FILING DATE: 1999-11-9

; NUMBER OF SEQ ID NOS: 469

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 412

; LENGTH: 433

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-711-164-412

Query Match 6.6%; Score 168; DB 4; Length 433;

Best Local Similarity 23.0%; Pred. No. 3e-08;
Matches 120; Conservative 59; Mismatches 176; Indels 166; Gaps 28;

QY 1 MGGSGDDDDLLALAVTKSSLLIVGAGTWTGTSTALHLARRGYTNVTVDLPYPVPSAISAGN 60

DB 7 LGHDMEDD-----CDIIIGAGTAGTACALCARAG-LSVLLLEAEIFG----- 50

QY 61 DVNKVISSGQYSNNKDEIEVNEILAEAFNGWKNDPLFKPYHYDHTGLMSACSGEGLDRL 120


```
QY 436 KWTNVQYRD 445
Db 434 -KWTQYTE 442

RESULT 6
US-09-252-991A-21043
; Sequence 21043, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21043
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21043

Query Match 5.8%; Score 146.5; DB 4; Length 448;
Best Local Similarity 23.7%; Pred. No. 5.1e-06;
Matches 98; Conservative 59; Mismatches 164; Indels 93; Gaps 23;

QY 18 SLLIIVGAGTGTSTALHARRGYTNVTLDPYPSAISAGNDVNVKVISSGOYSNNKDE 77
Db 51 TDVCVIGAGTGTSTALFLLENGF-KVTVLETAKVGFAS-GRNGGOIVNS--YSRDIDV 106

QY 78 IEVN-----EILAEAFNGKNDPLFKPYHYDTGL-----LMSACSQEGLDRLGVR 123
Db 107 IERTVGKREAQLGEMAFEGR---IIRERVYGTQCDLKGGVFAAFTEKQMDHLRAQ 163

QY 124 VRGGE--DPLVELTRPEQFKLAPEGVLQGDPPGKYGFPARGAGWAHARNALVAAARE 181
Db 164 KQLWERYGHNLQIMDAKRIR---EVVATDNYIG--GMLDMSG-GHIHPLNLALGEAAA 216

QY 182 AQRMGVKFTVGTQGRVVVTILFENNVDKGVAVTGDGKIWAERTFLCAGASAGOFLODFNQ 241
Db 217 VESLGGRIYEQSPATRI-----ERGASPVVHTPQGV---KAKPIVVAGNA---YLN---- 262

QY 242 LRPTAWTLVHIALKPEERALKYKNIP-----VI-----FNIERGFFPEDEERGEIKICDEHP 293
Db 263 -----GLVPELAA--KSMPCGTQVITTEPLNEELAHSLLPQD-----YCVEDC 303

QY 294 GYTNMVQASDGTWMSI-----PFSEKTIQPKAEATRVRRALLKETMPQLADRPFSFARICWCA 349
Db 304 NYLLDYRLSGDKRLIYGGGVYICARDPADIEALIRPKMLKTPQLKDVKIDFA---W-- 358

QY 350 DTANREPLIDRHQO-----VHSLVLCGASRGFKYLPISGNLIVDAMEGK 395
Db 359 -TGNFLLTSLRPQVGRIGNDIYISQ-----GCSHGVTYTHVAGKVLAEALRGQ 407

RESULT 7
US-09-883-096-5
; Sequence 5, Application US/09883096
; Patent No. 6680369
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Craven, Andrew
; APPLICANT: Yu, Ming
; APPLICANT: Sakowicz, Roman
; APPLICANT: Patel, Umesh A.
; APPLICANT: Davies, Katherine A.
; TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR USE
```

```
; FILE REFERENCE: 020552-001410US
; CURRENT APPLICATION NUMBER: US/09/883,096
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 09/594,655
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HsKip3a
; OTHER INFORMATION: fragment
; OTHER INFORMATION: Amino acid sequence of HsKip3a fragment used in
; OTHER INFORMATION: the ATPase assay (Figure 4).
US-09-883-096-5

Query Match 5.4%; Score 136.5; DB 4; Length 383;
Best Local Similarity 22.9%; Pred. No. 4.2e-05;
Matches 96; Conservative 44; Mismatches 163; Indels 117; Gaps 19;

QY 122 VVRPGEDPNLVELTRPEQFKLAPEGVL-----QGDFFPGWK----- 158
Db 12 VVRPPTPRELDSQRRP--VVQVVDERVLVFNPEEPDGGFPGLKGGTHDGPKKKGKDLT 69

QY 159 -----GYFARSAGAGWAHARNALVAAAREAQRMGVKFTVGTQGRVVTLIFENNVDVKA 211
Db 70 FVFDRVFGEAATQODVFQHTTHSVLDSFLQGYNCNVFAYCATGAGKTHMLGRE----- 123

QY 212 VTGDGKIMWAERTFLCAGASAGQ-----FLDFKNQLRPTAWTLVHIALKPEERAL 261
Db 124 --GDGIMVLTVELVRRLEARQQEKHFELISYQVYNE-----QIHDLLEPKGPLA 174

QY 262 YKNIPVIFNIERGEF-PFEDDEERGEIKIC-----DEHPGYTNMVQASDGTWMSIPPEK 313
Db 175 IREDDPKGVVQGLSPHQPASAEQLLEILTRGNRNRTQHTPTDANATSSRSHAIFQI--FVK 233

QY 314 TQIPKEAETRVRRALLKETMPQLADRPFSFARICWCADTANREPLIDRHPOYHSLVLCGCA 373
Db 234 QQDRVPEGLQAVQVAKNSL-----IDLASERASSTHAKGERLRREGANI 277

QY 374 SGRGFKYLSIGNLIIVDAMEGK--VPQKIHLELIKWNPDIAANRNWRDITLGRFGGPNRVM- 430
Db 278 N-RSLLALINVLNALADAKGRKTHVPEYR-----DSKLTLLKDSLG--GNCRTVMI 325

QY 431 -----DFHDVKETNVQYRDISK---LKG--ELEGLPINPILL-----RTCHHHHHH 472
Db 326 AAISPSLSLYEDT--YNTLKYADRAKEIRLKGNKSLGKGPINPILLGLDSTRTHHHHHH 383

RESULT 8
US-09-199-637A-367
; Sequence 367, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 367
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-199-637A-367

Query Match          5.4%; Score 135.5; DB 3; Length 378;
Best Local Similarity 24.3%; Pred. No. 5.1e-05;
Matches 102; Conservative 45; Mismatches 165; Indels 107; Gaps 19;

Qy 17 SSSLLIVGAGTGTSTALHARRGTYNTVLDPPVPSAISAGNDVNVKVISSGGYNNKD 76
Db 16 SRDVVVVGAGVIGLLTARELALAGLRVTLVERGESGREASWAGGIVSPLYPRYSP--- 72
Qy 77 EIEVNEILAEAFNGWKNDPLFKPYHYDHTGLLMSACSQEGLDRLGVRVRPGEENL--- 132
Db 73 -----AVTALAHSQD---FYP-----ALQORLLDETGL-----DPEVHTVG 106
Qy 133 ---VEL-----TRPEQFRKLAPGEGVLQDPPGKWKGYFARS--GAGWAHARNA 174
Db 107 LYLWLDLDDQTEALQWARKHTRP---LKEVPIEEAYAAVPGLGAGFQRAVYMSGVANVRNP 163
Qy 175 LVAAREAQRMGVKFTGTGQGRVVVTILFENNVDKGAVTGDKIWRARTFLCAGASAGQ 234
Db 164 RLARSLRASLQOQFANLHELHEQTEVRGWLDRDGRVGVATSRGEI-RGDKVLLAAGAWSGE 222
Qy 235 FLDFKQNLQRPNTAWTLVHIALKPEERALKYK-----NIPVFNIERGFFFPDBERGEIKICD 290
Db 223 L-----LKLGLLELPVVPVK-GOMILYKCAADFLPRMV-LAKGRYAIPRRD-GHILIGS 273
Qy 291 --BHPGYTNMVQSADGTMTSIPFEKTOIPEAEATRVRLALKETMPOLAD-----R 338
Db 274 TLEHSG-----FDKTPDDEALES-LRASAAELLPELADMQPVNAHWAGLR 316
Qy 339 PFSFARICWCADTANREFLIDRHPQYHSLVLCGCGSRGFKYLPISIGNLIIVDAMEGKVP 397
Db 317 PGSPEGIPY-----ICPVPGDGLWLTNGHYRNLGLVLAFLAPASCELLADLMSGREP 365

RESULT 9
US-09-252-991A-21423
; Sequence 21423, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21423
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-21423

Query Match          5.4%; Score 135.5; DB 4; Length 419;
Best Local Similarity 24.3%; Pred. No. 6.1e-05;
Matches 102; Conservative 45; Mismatches 165; Indels 107; Gaps 19;

Qy 17 SSSLLIVGAGTGTSTALHARRGTYNTVLDPPVPSAISAGNDVNVKVISSGGYNNKD 76
Db 57 SERVVGAGVIGLLTARELALAGLRVTLVERGESGREASWAGGIVSPLYPRYSP--- 113
Qy 77 EIEVNEILAEAFNGWKNDPLFKPYHYDHTGLLMSACSQEGLDRLGVRVRPGEENL--- 132
Db 114 -----AVTALAHSQD---FYP-----ALQORLLDETGL-----DPEVHTVG 147
```

```
Qy 133 ---VEL-----TRPEQFRKLAPGEGVLQDPPGKWKGYFARS--GAGWAHARNA 174
Db 148 LYLWLDLDDQTEALQWARKHTRP---LKEVPIEEAYAAVPGLGAGFQRAVYMSGVANVRNP 204
Qy 175 LVAAREAQRMGVKFTGTGQGRVVVTILFENNVDKGAVTGDKIWRARTFLCAGASAGQ 234
Db 205 RLARSLRASLQOQFANLHELHEQTEVRGWLDRDGRVGVATSRGEI-RGDKVLLAAGAWSGE 263
Qy 235 FLDFKQNLQRPNTAWTLVHIALKPEERALKYK-----NIPVFNIERGFFFPDBERGEIKICD 290
Db 264 L-----LKLGLLELPVVPVK-GOMILYKCAADFLPRMV-LAKGRYAIPRRD-GHILIGS 314
Qy 291 --BHPGYTNMVQSADGTMTSIPFEKTOIPEAEATRVRLALKETMPOLAD-----R 338
Db 315 TLEHSG-----FDKTPDDEALES-LRASAAELLPELADMQPVNAHWAGLR 357
Qy 339 PFSFARICWCADTANREFLIDRHPQYHSLVLCGCGSRGFKYLPISIGNLIIVDAMEGKVP 397
Db 358 PGSPEGIPY-----ICPVPGDGLWLTNGHYRNLGLVLAFLAPASCELLADLMSGREP 406

RESULT 10
US-09-252-991A-24053
; Sequence 24053, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24053
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-24053

Query Match          5.3%; Score 134.5; DB 4; Length 484;
Best Local Similarity 22.4%; Pred. No. 9.7e-05;
Matches 103; Conservative 70; Mismatches 185; Indels 101; Gaps 23;

Qy 12 LAVTKSSLLIVGAGTGTSTALHARRGTYNTVLDPPVPSAIS-----AGND 61
Db 77 LAGEEQADVCVGGSGFSGVNTALELAQRGP-SVLLLEAHRIGWASGRNGQLIRGVGHD 135
Qy 62 VNK---VISSGQYNNKDEIEVNEILAEAFNGWKNDPLFKPYHYDHTGLLMSACSQEGLD 118
Db 136 VEQFFPVIGA-----DGVKALKMLCGLAVEIVR--RVEYAIADCIDLNGYCD----- 181
Qy 119 RLGRVVRPG-----EDPNLVELTRPEQFRKLAP-----EGVLQDPPGKWKGYFARSAGW 168
Db 182 ---LANKPGDYQCFREDMEBELQALGYRHEMLVPAEAMRSVVGSD--RYVGLVDMGSGH 236
Qy 169 AHARNALVAAAREAQRMGVKFTGTGQGRVVVTILFENNVDKGAVTGDKIWRARTFLCA 228
Db 237 LHPNLNLVLGEEAAAQSLGVRLEPERSP-----VTRIDYGAEVQ-VHTATGKV-RAKT--LVL 288
Qy 229 GASAGQFLDFKQNLQRP-----TAWTLVHIALKP-EERALKYKNIPIVFNIERGFFFPD 280
Db 289 GCNA-----YMNDLNPLGKGLVPAGSYV-IAEPLDEQLARQLLP----- 328
Qy 281 EERGEIKICDEHPGYTNMVQSADGTW--SIPPEKTQIPKEAETRVRLALKETMPOLADR 338
Db 329 ---QNMVACDQORVALDYRLSADNRLFGGACHYSGRDPDSIAAYMRPKMLEVFPQLANV 385
Qy 339 PFSFARICWCADTANREFLIDRHPQYHSLVLCGCGSRGFKYLPISIGNLIIVDAMEG--- 394
```

Db 386 RIDYQGMIGIGANRLPQIGRLPGQPNVYFAQYSGHGVNATHLAGQLLAEAGGQSD 445
QY 395 -----KVPQKIH---ELIKWNPDIANRW---RDTLG 421
Db 446 GFDLFAKVPHITPGGKLLR-SPLLALGAWYRLKEKLG 483
RESULT 11
US-09-252-991A-30872
; Sequence 30872, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30872
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30872

Query Match 4.9%; Score 125; DB 4; Length 652;
Best Local Similarity 21.8%; Pred. No. 0.0015;
Matches 92; Conservative 54; Mismatches 162; Indels 114; Gaps 16;
QY 17 SSSLLIVGAGTGTSTALHARRGYTNVTVDLP-YPVPSAISAGNDVNVKVISSGQYNNK 75
Db 283 NAEVIVVGAGIVGSACAHELARRG-LDVLVDSRRGGATAVGMGLV-----AMDDNP 334
QY 76 DETEVNEILAEAFNGWKNPDKPKPYVHDGGLMSAC-----SQEGLDRLGVVR 125
Db 335 AELALSD-YSIQAWRTWAADLPEDCAYRNGCTLWLADAELAEAEKQALLAAGV--- 390
QY 126 PGEDPNLVELTRPEQFKLAPEGVLQDFFGKGYFARGAGWAHARNALVAAAEARQM 185
Db 391 -----ACEMLDAARLDLEP---VLR---PGLAGALKVPDGDILYAPN---AARWLLERA 436
QY 186 GVKFVTGTPGQVVTLLFENNVDKGA---VTGDGKIWRAERTFLCAGASAGQFLDFKNQLR 243
Db 437 G-----PRLRLHAEVSEVDGSRRLADGRWLSAEALVLANGIHAGEL----- 479
QY 244 PTAWTLVHIALKPEERALYKNIPVIFNIERGFFPEDEERGEIKICDEHP----- 293
Db 480 -----CAELPIRP-----KKGHLLITDYPGTLRHQLVEL 509
QY 294 GYTNMVQASDGTWM-----SIPFEKTQIPKEAETRVALLK--ETWQOL 335
Db 510 GYVSSAHASSGTSVAENAPRPTGQVFLGSSRQFDTLDPQVEGVLARMRLRLLDVLPGL 569
QY 336 ADPRFPFARICWADTANREFLDHHPQYHSLVLCGASGRGFKYLPISGNLIVDAMEGK 395
Db 570 AGLNAIRAWTGFRAATPDGLLGEHPAQPGFLWLVGHEGLVTTAPGSARLLAAQLFGE 629
QY 396 VP 397
Db 630 TP 631

RESULT 12
US-09-489-039A-12890
; Sequence 12890, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12890
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12890
Query Match 4.9%; Score 124.5; DB 4; Length 476;
Best Local Similarity 22.6%; Pred. No. 0.001;
Matches 94; Conservative 48; Mismatches 153; Indels 121; Gaps 21;
QY 11 ALAVTKSSLLIVGAGTGTSTALHARRGYTNVTVDLPY-----PVPISAISA 58
Db 66 ALREDIEADVIIIGGFGSINTALEAEQGITNVVVLRLHGLGYGTGRNGGQVMAGI-- 123
QY 59 GNDVNVKVISSGQYNNKDEIE-----VNEILAEAFNGWKNPDKPKPYVHDGGLMSA 111
Db 124 GHDIKAV-----KKHVGKGLTLEFKIANLGAGIIRIRIKYNIDADFVPGY---GYL--A 174
QY 112 CSQEGDLRLG-----VRVRPGEDPNLVELTRPEQFKLAPEGVLQDFFGKGYFARG 165
Db 175 YNQRQLKTLRQWEKEPKAATPDEE---IELYTGKEVQVVGSEV-----YCGALKHMG 224
QY 166 AGWAHARNALVAAAEARQMGVKFVTGTP-----QGRVVTLLFENNVDKGAVTGDGK 217
Db 225 GGQIHSNMLLGSAAQAHSLGVKIFESSPVVEVNYGKQVRVTAM-----GSVKAACL 277
QY 218 IWRPAERTFLCAGASAGQFLDFKNQLRPTAW---TLV-----HIALKPEERALYKNIPVIFNI 271
Db 278 LWACD-----SFLNNMEPEIYNKTLVTYTSQVSTEPSDELIERISPL--- 320
QY 272 ERGFFPEDEERGEIKICDEHPGYTNMVQASDGTMMSSIPFEKTQIPKEAETRVALLKET 331
Db 321 -RGAFSDIRPVINYVTRT-----NRLJFGSATRF---VEYT--PNDFAWNRTLLAEV 369
QY 332 MPOLADRPFSFARICWADTANREFLDHHPQYHSLVLCGASGRGFKYLPISGNL 387
Db 370 FPYLRDVKIDFA---MGGPMA-----CSAN-----LFPQIGTL 399

RESULT 13
US-09-684-405-2
; Sequence 2, Application US/09684405
; Patent No. 6770477
; GENERAL INFORMATION:
; APPLICANT: Dennis J. Slamon
; APPLICANT: Juliana J. Oh
; TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED GENES
; TITLE OF INVENTION: ASSOCIATED WITH HER-2/NEU OVEREXPRESSION
; FILE REFERENCE: 30448.79USU1
; CURRENT APPLICATION NUMBER: US/09/684,405
; CURRENT FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/157,923
; PRIOR FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-684-405-2
Query Match 4.9%; Score 123.5; DB 4; Length 486;
Best Local Similarity 20.9%; Pred. No. 0.0013;
Matches 94; Conservative 53; Mismatches 179; Indels 123; Gaps 20;
QY 18 SSSLLIVGAGTGTSTALH-----ARRGYTNVTVDLPYVPISAISAGNDVNVK----- 65

Db 64 SDVVVGGVGLSVAVLWKLSESGAIRVLVVERDHTYSQASTGLSVGGICQOFSLPE 123
Qy 66 -----ISSQYNNKNDIEVNEILAEBAFGWKNDFLPKPYHDTGLLMSACSQEGLD-R 119
Db 124 NIQLSLFSASFLRN-----INEYLAV-----VDAPPLDLRFNPSGYLLLLASEKDAAME 172
Qy 120 LGVRVRGEPDNLVELTRPEQFRKLAP-----EGVLQDFFGKGYFARSAGNAHARNAL 175
Db 173 SNKVQVQEGAK-VSLMSPDQLRNKFPWINTGVALAS-----YGMEDGWFDPWCLL 224
Qy 176 VAAAREARMGVKEVTG-----TPQGRVVTL-----IPENNVDVG 210
Db 225 QGLRRKVQSLGVLEFCQEVTRFVSSQSRLTTDDKAVLKRHHVHVMDRSLEYQVPEC 284
Qy 211 A-VTGDGKIWAERTFLCAGASAGQFLDFKNQLRPTAWTLVHIALKPEERALKYKNIPVIF 269
Db 285 AIVINAAGAWSAQIAAL-AGVGECP-----PGTLQGTKLPEPRKRVY----- 327
Qy 270 NIERGPFPEDEERGEIKICDEHPGY-TNMVQSADGTW-----MSIPFEKTOIPKEATR 323
Db 328 -----VMHCPQSGLETPLVADTSGAYFRREGLSNYLGGRSPTEQEEP 371
Qy 324 VRALLK-----ETWPLQADRPFSFARI-----CWCA-----DTANREFLIDRHPQVHSL 367
Db 372 DPNLEVDHDFQDKVWPHLALRVAPETIKVQSAWAGYVDYNTFDQGVVGHPLVVM 431
Qy 368 VLGGASGRGFKYLPISIGNLIVD-AMEGK 395
Db 432 YFATGFSCHGLQAPGIGRAVENVLKGR 460

RESULT 14

US-09-213-888-25
; Sequence 25, Application US/09213888A
; Patent No. 6638731
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Li, Jinhe
; APPLICANT: Pauley, Adele M.
; TITLE OF INVENTION: Human Sel-10 Polypeptides and Polynucleotides that
; FILE REFERENCE: 6142
; CURRENT APPLICATION NUMBER: US/09/213.888A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: V5HIS tagged
; OTHER INFORMATION: homo sapien
US-09-213-888-25

Query Match 4.8%; Score 121.5; DB 4; Length 669;
Best Local Similarity 51.9%; Pred. No. 0.0035;
Matches 27; Conservative 6; Mismatches 10; Indels 9; Gaps 4;

Qy 429 VMDPFH-DVKEWTN--VOYRDISKLKGEGLPIPNPLL-----RTGHHHHHH 472
Db 619 VLDFDVMKFCRYPAQWRPLES-RGPFEGKPIPNPLLGLDSTRTGHHHHHH 669

RESULT 15

US-09-328-877D-25
; Sequence 25, Application US/09328877D
; Patent No. 6730778
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Li, Jinhe

; APPLICANT: Pauley, Adele M.
; APPLICANT: Pharmacia & Upjohn Company
; TITLE OF INVENTION: Human Sel-10 Polypeptides and Polynucleotides that
; FILE REFERENCE: 6142
; CURRENT APPLICATION NUMBER: US/09/328.877D
; CURRENT FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: V5HIS tagged
; OTHER INFORMATION: homo sapien
US-09-328-877D-25

Query Match 4.8%; Score 121.5; DB 4; Length 669;
Best Local Similarity 51.9%; Pred. No. 0.0035;
Matches 27; Conservative 6; Mismatches 10; Indels 9; Gaps 4;

Qy 429 VMDPFH-DVKEWTN--VOYRDISKLKGEGLPIPNPLL-----RTGHHHHHH 472
Db 619 VLDFDVMKFCRYPAQWRPLES-RGPFEGKPIPNPLLGLDSTRTGHHHHHH 669

Search completed: May 27, 2005, 13:50:01
Job time : 37.2303 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:34:21 ; Search time 36.2303 Seconds
(without alignments)
1253.491 Million cell updates/sec

Title: US-10-622-893A-5
Perfect score: 2530
Sequence: 1 MGSGDDDLALAVTKSSSL.....LEGLPINPLRTGHHHHH 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	488.5	19.3	412	2 T40295	fructosyl amine -
2	465	18.4	433	2 T37605	probable fructosyl
3	192	7.6	390	2 T39975	sarcosine oxidase
4	181	7.2	389	2 JS0671	sarcosine oxidase
5	177.5	7.0	390	2 JC7256	L-pipecolate oxida
6	170	6.7	389	2 T44248	sarcosine oxidase
7	168	6.6	433	1 B65058	fixC protein homol
8	155	6.1	387	2 JU0461	sarcosine oxidase
9	151.5	6.0	433	2 E91081	hypothetical prote
10	151.5	6.0	433	2 F85926	hypothetical prote
11	148	5.8	380	2 AB3160	santhopine deamina
12	146.5	5.8	427	2 G83298	conserved hypothet
13	142.5	5.6	857	2 S16133	dimethylglycine de
14	139.5	5.5	364	2 B83078	probable D-amino a
15	138	5.5	376	2 A83905	hypothetical prote
16	137.5	5.4	372	2 A85668	sarcosine oxidase-
17	137.5	5.4	372	2 E90808	sarcosine oxidase-
18	136.5	5.4	372	2 G53371	probable sarcosine
19	136.5	5.4	382	2 J75122	sarcosine oxidase,
20	136.5	5.4	416	2 AD3075	sarcosine oxidase
21	136.5	5.4	416	2 E98211	sarcosine oxidase
22	136.5	5.4	825	2 F95963	probable dehydroge
23	135.5	5.4	372	2 F90159	sarcosine oxidase,
24	132	5.2	371	2 F83487	hypothetical prote
25	129.5	5.1	439	2 E82983	probable oxidoredu
26	128.5	5.1	382	2 F71008	probable sarcosine
27	127.5	5.0	433	2 A10882	probable oxidoredu
28	127	5.0	372	2 A10637	probable sarcosine
29	127	5.0	372	2 A70019	opine catabolism h

30	126	5.0	652	2 AH2245	thiamin biosynthes
31	125.5	5.0	417	2 G98200	sarcosine oxidase
32	125.5	5.0	417	2 AB3086	sarcosine oxidase
33	125.5	5.0	442	2 D97382	hypothetical prote
34	125.5	5.0	442	2 AC2600	oxidoreductase ord
35	123.5	4.9	437	2 T45533	agaE protein (impo
36	116.5	4.6	452	2 G95356	probable oxidoredu
37	114	4.5	396	2 AF2835	sarcosine oxidase
38	114	4.5	413	2 A97613	sarcosine oxidase
39	113	4.5	420	2 AC3284	D-amino-acid dehyd
40	113	4.5	703	2 T15503	hypothetical prote
41	112.5	4.4	629	1 BVECOA	glucose inhibited
42	112.5	4.4	629	2 C91214	glucose-inhibited
43	112.5	4.4	629	2 D86060	glucose-inhibited
44	112	4.4	405	2 I40887	sarcosine oxidase
45	111.5	4.4	441	2 F98306	agaE protein (impo

ALIGNMENTS

RESULT 1

T40295
fructosyl amine - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T40295
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Hilbert, H.; Duesterhoeft, A.
submitted to the EMBL Data Library, March 1998
A;Reference number: Z21918
A;Accession: T40295
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-412 <WOO>
A;Cross-references: UNIPROT:O43029; EMBL:AL022071; PIDN:CAA17815.1; GSPDB:GN00067; SPDB:
A;Experimental source: strain 972h-; cosmid c354
C;Genetics:
A;Gene: SPDB:SPBC354.15
A;Map position: 2

Query Match	19.3%	Score 488.5;	DB 2;	Length 412;
Best Local Similarity	30.1%	Pred. No. 1.3e-30;		
Matches 130;	Conservative	78;	Mismatches 183;	Indels 41; Gaps 12;
Qy	14	VTKSSLLIVGAGTWTGTSTALHARGYTNVTLDYPVPSAISAGNDVNKVISSGQYSN	73	
Db	1	MVKNTSVIIIVGAGVGLSAAELTKRGYTIKILDRAPPVPVIGDSSVDANRIIRS-DYAD	59	
Qy	74	NKDEIFVNEILAEAFNGKNDPLFKPYHYDHTGLLSA-----CSQEGDLRLGVVR	125	
Db	60	-----AVYCSMGIDALEEWRNPLFKQFYGSGLMFVGRDNVEYRDMLENLTKMGV---	111	
Qy	126	PGBDPLNVLTRPEPRKLAPEGVLQDFFQWKGYFARSGAGWAHARNALVAAAREAQR	185	
Db	112	-----SAKQOTTEELRLKLPKI--GELNDGEAGYANFSSGWNARQSVKSVVNYLAHA	164	
Qy	186	GVKFTVTGQGVVVTLIFENNVDKGAVTGDKIWRABRTFLCAGASAGQFL-DFKNQLRP	244	
Db	165	GVSFISG-PEGTVLEELITEENVVKGVRTTG-AVMAEKLI FATGAWTASLLPNHTRFLA	222	
Qy	245	TAMTLVHIALKPERALYKNIPVIFNIERGFPEPDEERGEIKCDEHPGTYNMVQSADG	304	
Db	223	TGQPVAVIKLTPEYIRFLTNPNVYLDFTGTGYIPTPDGVLKPFARHGYGFTRMQLKSG	282	
Qy	305	TWMSIPEKQTQI---PKEATRVALLKETM-POLADRPSPFARICWCADTANREFLID	359	
Db	283	KVSVPPKPLDVPILPKAEALDRLRLQRTYGEISQRPFYKTRI CYTTDTAAEFVFD	342	
Qy	360	RHPQYHSLVLCGASGRGFKYLPISIGNLIVDAMEGKVPQKIHILIKW---NPDIARNRW	416	
Db	343	YHPDENLNVCTGSGGHGFFPFLIGKYSIGCMFRELEELPKKWRKKNLEFAALDHS	402	
Qy	417	RDTLGRFGGPNR	428	

Db 403 R-----AGFSR 408

RESULT 2

T37605

probable fructosyl amino acid oxidase - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C;Accession: T37605

R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Saunders, D.; Harris, D.

submitted to the EMBL Data Library, October 1999

A;Reference number: Z21729

A;Accession: T37605

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-433 <MCD>

A;Cross-references: UNIPROT:Q9UTW9; EMBL:AL132667; PIDN:CAB59618.1; GSPDB:GN00066; SPDB:

A;Experimental source: strain 972h-; cosmid c139

C;Genetics:

A;Gene: SPDB:SPAC139.04C

A;Map position: 1

Query Match 18.4%; Score 465; DB 2; Length 433;

Best Local Similarity 30.1%; Pred. No. 1e-28;

Matches 130; Conservative 80; Mismatches 174; Indels 48; Gaps 13;

QY 17 SSSLLVAGTGTSTALHILAR-RGYTNVTLDPYVPVPSAISAGNDVNKVISSGQYNNK 75

Db 2 SRTIVVGGVGLSTAVELAKNHSFDNIITDAEPVPSMSAANDINKIVRP-EYADLK 60

QY 76 DETEVNEILAEAFNGWKNDPLFKPYVYHDTGLMSACSQEGLDRLGVVRVPGF----- 128

Db 61 -----YMKLALEAMEKWRNDPELSSVYFEGRLST-----ISKDYRAEFDEVAQRNL 109

QY 129 -----DPNLVELTPEQPKLAP-----EGVLQDFFPGWKGYFARSGAGWAHARNALVAAR 180

Db 110 KLLGDSALINLSSEERKKYKPSLFSNPLRSDMQA-----VNEHAGYANSAASLKLEL 165

QY 181 EAQRMGVKFTYGTPOGR-----VV-----TLIFENDVKGAVTGDKIWRARTLCLGASA 232

Db 166 KARELGVEFVFG-KAGFKFKFVNNHSETDIDKDNHVSQVQTEGDIYHADIILLAVGAYL 224

QY 233 GQFLDFKNQRLPTAWTLVHIALKPEERALKNPVFNIERGFFFPDEERGEIKICDEH 292

Db 225 NAYLNTSHRVCAGLPAVHQLTDEEPKTYKNWPIIFDPCAYAFPPYPTVKLIKLASTG 284

QY 293 PGYTNMVQS---ADGTMMSIPF---EKTQIPKAEATRVALLKETMPQLADRPSPFARIC 346

Db 285 YEYVCNVETDYDENSKVVSIPHSGPSKSLPKYALIQMRFLDTFLPDLADRSILINTKMC 344

QY 347 WCADTANREFLIDBHPQYHSLVLCGASGRGFKYLPISGNLIVDAMEGKVPQKIHILIKW 406

Db 345 WISDTEDANFLIDKVPQDFNVFVANGDSGHAFKFLPNIGRYIAQRIILGDLSEWKKDARW 404

QY 407 NPDIANR-NWR 417

Db 405 REDDKASELKWR 416

RESULT 3

I39975

sarcosine oxidase (EC 1.5.3.1) - Bacillus sp.

C;Species: Bacillus sp.

C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 09-Jul-2004

C;Accession: I39975

R;Suzuki, K.; Imamura, S.; Sugiyama, M.

J. Ferment. Bioceng. 77, 231-234, 1994

A;Title: Cloning, sequencing, overexpression in Escherihia coil of a sarcosine oxidase-e

A;Reference number: I39975

A;Accession: I39975

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-390 <RES>

A;Cross-references: UNIPROT:P40859; GB:D16521; NID:G984787; PIDN:BAA03967.1; PID:G984788

C;Genetics:

A;Gene: sox

C;Superfamily: sarcosine oxidase

C;Keywords: oxidoreductase

Query Match 7.6%; Score 192; DB 2; Length 390;

Best Local Similarity 23.2%; Pred. No. 2.7e-07;

Matches 97; Conservative 68; Mismatches 162; Indels 92; Gaps 22;

QY 20 LLIVGAGTGTSTALHILARGYTNVTLDPYVPVPSAISAGNDVNKVISSGQYNNKDEIE 79

Db 7 VIVWGAGSMGMAAGYQLAKQGVKTLV-DAPDPHTNGSHHGDTLRIHA-YGEGREYVP 64

QY 80 VNEILAEAFNGWKNDPLFKPYVH-----DTGLMSACSQEGLDRLGVVRVPGDPLNVEL 135

Db 65 ----LALRQELWYE--LEKETHHKIFTKTGLV-----FGPK--GESAFVAET 105

QY 136 TRPEQFKLAPEGVLOGD-----FPGW---KGYYA--RSAGWAHARNALVAAREAORM 185

Db 106 MEAAKEHSLTVD-LLEGDEINKRWPGITVPENYNAIFEPNSGVLFSENCIRAYRELAEAR 164

QY 186 GVKFVTGTPOGRVVTLLIPENDVKGAVTGDKIWRARTFLCAGASAGQFLDFKNQLRPT 245

Db 165 GAKVLTHT---RVEDFDISPDVSK-IETANGS--YTADKLIVSGAWNSKLLSKLN----- 214

QY 246 AWTLVHIALKPEERALKNPVFNIERGFFFPDEERGEIKICDEHPGVTNNVQSADGT 305

Db 215 ----LDLPLQ-----YRQV-----GPFESDESKYSNDI--DFPGF--MVEVPNGI 253

QY 306 MMSIPF-----EKTQIPKAEATRVALLKETMPQLADRPFSF 342

Db 254 YYGFPFSGGGLKGLGYHTFGQKIDPTINKRFYVPEDESNLRAFLLEYMPG-ANGELKR 312

QY 343 ARICWCADTANREFLIDRHPQYHSLVLCGASGRGFKYLPISGNLIVD-AMEGKVPQKI 400

Db 313 GAVCMYTKLDEHFIILHPEHSNVVIAAGFSGHGFKFSSGVEVLSQLALTGKTBDHI 371

RESULT 4

JS0671

sarcosine oxidase (EC 1.5.3.1) precursor - Streptomyces sp.

C;Species: Streptomyces sp.

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 20-Jun-2000

C;Accession: JS0671; PS0345

R;Suzuki, K.; Ogishima, M.; Sugiyama, M.; Inouye, Y.; Nakamura, S.;

Bioeci. Biotechnol. Biochem. 56, 432-436, 1992

A;Title: Molecular cloning and expression of a Streptomyces sarcosine oxidase gene in St

A;Reference number: JS0671; MUID:92330009; PMID:1368326

A;Accession: JS0671

A;Molecule type: DNA

A;Residues: 1-389 <SUZ>

A;Cross-references: GB:D10623; NID:G217023; PIDN:BAA01473.1; PID:G217024

A;Experimental source: strain KB210-8SY

A;Accession: PS0345

A;Molecule type: protein

A;Residues: 2-31;122-143;230-259;268-283 <SUZ1>

A;Note: the authors translated the initiation codon GTG for residue 1 as Val

C;Comment: This enzyme catalyzes the oxidation of sarcosine to form formaldehyde, glycine

C;Superfamily: oxidoreductase

C;Keywords: oxidoreductase

P;2-389/Product: sarcosine oxidase #status experimental <SAR>

P;11-16/Region: nucleotide binding #status predicted

Query Match 7.2%; Score 181; DB 2; Length 389;

Best Local Similarity 22.9%; Pred. No. 2e-06;

Matches 100; Conservative 59; Mismatches 150; Indels 128; Gaps 24;

QY 20 LLIVGAGTGTSTALHILARGYTNVTLDPYVPVPSAISAGNDVNKVISSGQYNNKDEIE 79

Db 7 VIVVGLGGMGSAAHHLISARG-ARVIGLEKF-----GPFVNRGSSHG 47

Qy 80 VNEILAEAFNGKNDPLFKPYVHTDTGLLMSACS-QSGLDRL-----GVRVPEGE 128
Db 48 GSRITRQSYF-----EDPAYVP-----LLRAYELYELERATGRNVTATLCGGVNAAGPDD 97
Qy 129 DPNLVELTRP-----EQFRKLAPEGVLQDFFPGWKGYFARSAGWAHA 171
Db 98 SRTVSGSLRSATWDLAHEMLDAKEIRRRPPTLAPDDDEVALF-----EAKAGLLRP 149
Qy 172 RNALVAAREARQMGVFKVTGTPQGRVVTLLIFENNVDKAVTGDG-KIWAERTFLCAGA 230
Db 150 ENMVAALHQLATROGAE-----LRFEEPLVRWBPYRDGVRVHTGENTY----- 192
Qy 231 SAGQFLDFKNQLRPTANTLVHIALKPEERALKYKNIPIVFNLERG--FFEPDDEBERGIKI 288
Db 193 TAGQLV-----ICGAWA-----PO-----LLADIGVPIITVERQIMWFQPKGTGPF-V 236
Qy 289 CDEHPGTNNVQSADGTMW-SIP-----FEKTO--IPKEAET-----RVRAL 327
Db 237 PERHPVY--IWEDADGVQVYGFPAIDGPEKGAQVAFPRKGOHTTPTETIDRTVHAHEVRAM 294
Qy 328 LKETMPQLADRPFSF--ARICWCADTANREFLIDRHPQY-HSLVGGCAGSRGPKYLPISI 384
Db 295 ADHMSALIPDLGFTFLKAATCMYNTPDHEFVIARHPAHPESVTVACGFGSHGKFPVAV 354
Qy 385 GNLIWD-AMEGKVPQKI 400
Db 355 GEILADIALTGATAHPI 371

RESULT 5
JC7256

L-pipecolate oxidase (EC 1.-.-.-) - human
C;Species: Homo sapiens (man)
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C;Accession: JC7256
R;Jlalt, L.; de Kromme, I.; Oostheim, W.; Wanders, R.J.A.
Biochem. Biophys. Res. Commun. 270, 1101-1105, 2000
A;Title: Molecular cloning and expression of human L-pipecolate oxidase.
A;Reference number: JC7256
A;Accession: JC7256
A;Molecule type: mRNA
A;Residues: 1-390 <JL>
A;Cross-references: UNIPROT:Q9P029
C;Comment: This enzyme, a peroxisomal enzyme and a flavoprotein, which converts L-pipecolate to pipecolate.
C;Superspecies: UNIPROT:Q9P029
C;Keywords: P4D; flavoprotein; oxidoreductase

Query Match 7.0%; Score 177.5; DB 2; Length 390;
Best Local Similarity 21.7%; Pred. No. 3.7e-06;
Matches 86; Conservative 66; Mismatches 174; Indels 71; Gaps 16;
Qy 21 LIVGAGTGTSTALHLARRGTYNTVLDPVPSAISAGNDVNKVISSGOYNNKDBIEV 80
Db 11 IVIGAGLQGCFTAYHLAKH-RKRILLLEQFLPHSRGSSHCQSRIIRKAYLE-----DF 63
Qy 81 NEILAEAFNGKNDPLFKPYVHTDTGLLMSACSQEGDLRLGVRVRPGEDPNLVELTRPEQ 140
Db 64 YTRMHCECYQIWAQ-----LEHEAGTQLH--RQTGLLLGMKENQELKTIQANLSRQV 115
Qy 141 FRKLAPEGVLQDFF-----GWKGYFARSAGWAHARNALVA---AARE---AQRMGVK 188
Db 116 EHQCLSSEELKQRPFNIRLPRGEVGLLDNSG-GVIYAYKALRALQDAIROLGGIVRDGEK 174
Qy 189 FVTGTPQGRVVTLLIFENNVDKAVTGDGKIWAERTFLCAGASAGQFLDFKNQLRPTAWT 248
Db 175 VVEINP-GLLV-----VKTTSRSTQAKSLVITAGWTNQL-----LRPLGIE 216
Qy 249 LVHIALKPEERALKYKNIPIVFNIERGF-----FFEPDDEBERGIKICDEHPG 294
Db 217 MPLQTLAINCYWREMPGSGVGSQAAPFCFLWGLCPHHIYGLTGYEPLGMKVSYYHGN 276
Qy 295 YTNMVQSADGTMMSIPPEKTIQIPKEATRV-RALLKETMPQLADRPFSFARICWCADTAN 353

Db 277 H-----ADPERDCPTARTDI---GDVQLSSFVRDHLDPKPEP-AVIESCMYTNTPD 326
Qy 354 REFLLDRHPQYHSLVLCGASGRGFKYLPISIGNLIWD 390
Db 327 EQFLDRHPKYDNIVIGAGFGHGFKLAPVVGKILYE 363
RESULT 6
T44248
sarcosine oxidase (EC 1.5.3.1) [imported] - Arthrobacter sp. (strain TE1826)
C;Species: Arthrobacter sp.
A;Variety: strain TE1826
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 12-Jul-2004
C;Accession: T44248
R;Nishiya, Y.; Toda, A.; Imanaka, T.
Mol. Gen. Genet. 257, 581-586, 1998
A;Title: Gene cluster for creatinine degradation in Arthrobacter sp. TE1826.
A;Reference number: 222735; MUID:98223334; PMID:9563845
A;Accession: T44248
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-389 <NIS>
A;Cross-references: EMBL:AB007122; NID:g3116219; PIDN:BAA25926.1; PID:g3116221
A;Experimental source: strain TE1826
C;Keywords: oxidoreductase
Query Match 6.7%; Score 170; DB 2; Length 389;
Best Local Similarity 21.1%; Pred. No. 1.4e-05;
Matches 91; Conservative 69; Mismatches 171; Indels 100; Gaps 20;
Qy 12 LAVTKSSLLIVGAGTGTSTALHLARRGTYNTVLDPVPSAISAGNDVNKVISSGOY 71
Db 1 MSIKQDYDVIVVGASGMAAGYYLSKQGVKTLIV-DSFDPPHTNGSHHGDTIRHA-Y 58
Qy 72 SNNKDBIEVNEILAEAFNGKNDPLFKPYVHTDTGLL-----MSACSOBGL 117
Db 59 GEGREYVFP-ALRAQELWYLEKETHHK-IFTKTGLVLFVGPKGEAPFVAETMEAKHSL 116
Qy 118 DRLGVRVRPGEDPNLVELTRPEQPRKLAPEGVLQDFFPGWKGYFARSAGWAHARNALVA 177
Db 117 D-----VDLLESGSEINKRWP-GVTYPE--NYNAIFEKN-SGVLFSENCIRA 158
Qy 178 AAREARQMGVFKVTGTPQGRVVTLLIFENNVDK-----GAVTGDGKIWAERTFLCAGASA 232
Db 159 YRELAEEANGAKVLTYP---VEDPEIAEDFKIQTAYGSET-----ASKLIYVMGAWN 208
Qy 233 GQFLDFKNQLRPTANTLVHIALKPEERALKYKNIPIVFNIERGFFEPDDEERGE----- 285
Db 209 SKLLSKLN-----IEIPLQP-----YRQVV-----GFFECDEKKYSNTHGYPA 246
Qy 286 --IKI-----CDEHPGTNNVQSADGTMMSIPPEKTIQIPKEATRVRLALKE 330
Db 247 FMVEVPTGIYVGPSPFGCGGLKIGYHTYGGKIDPDTINREF---GIYPEDSEGNIRKFL 303
Qy 331 TMPQLADRPFSFARICWCADTANREFLIDRHPQYHSLVLCGASGRGFKYLPISIGNLIWD 390
Db 304 YMPG-ATGELKSGAVCMYTKTPTDEHFVIDLHPQSNVAIAAGFGSHGKFPSSVVGETLSQ 362
Qy 391 -AMEGKVPQKI 400
Db 363 LAVTGKTEHDI 373

RESULT 7
B65058

fixC protein homolog b2766 - Escherichia coli (strain K-12)
N;Contains: probable quinone reductase (EC 1.6.5.-)
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002
C;Accession: B65058
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-427 <STO>
A:CROSS-references: UNIPROT.Q91067; GB:AE004705; GB:AE004851; PIDN:AAG0616
A:Experimental source: strain PAOL
C:Genetics:
A:Gene: FA2776
C:Superfamily: hypothetical protein HI0499

Query Match 5.8%; Score 146.5; DB 2; Length 427;
Best Local Similarity 23.7%; Pred. No. 0.0012;
Matches 98; Conservative 59; Mismatches 164; Indels 93; Gaps 23;

QY 18 SSLIVGAGTGTSTALHLARRGYNTVLDPPVPSPSAISAGNDVNKVKISSGOYSNNKDE 77
: : : ||| ||| ||| :
Db 30 TDVCVICAGVTGLSTALFLLENG-F-KVTULEAAKVGFAS-GRNGGGQIVNS--YSRDIDV 85
: : : ||| ||| ||| :
QY 78 IEVN-----EILAEAFNGKNKDPLFKPYHDTGL-----LMSACSQEGDLRLGV 123
: : : ||| ||| ||| :
Db 86 IERTVGKREQLLGEMAFEGR---IIRVARVYGIQCDLKOGGVFAAFTEQMDDLRAQ 142
: : : ||| ||| ||| :
QY 124 VRPEE--DPNLVELTRPEQFRKLPAEVLQDPPGMKGYPARGAGWAHARNALVAARE 181
: : : | :
Db 143 KQLWERYGHNOLEIMDKRIR----EVATDNVIG--GMLDSG-GHIHPLNALGEAAA 195
: : : | :
QY 182 AQRMGVFVTGTPQGRVTVLIFENNVDKGAVTGDGKIWRARTFLCAGASAGQFLDFKNQ 241
: : : | :
Db 196 VESLGGRIYEQSPATRI-----ERGASPVVHTPOGV---KAKFIVVAGNA--YLN--- 241
: : : | :
QY 242 LRPTAWTLVHIALKPERALRYKNIP-----VI-----FNIERGFFEPDEERGETIKIDEHP 293
: : : | :
Db 242 -----GLVP ELAA--KSMPCGTQITTEPLENEELAHSLLPQD-----YCVEDC 282
: : : | :
QY 294 GYTNNVQSADGTWMSI-----PFKEQTQPKEAETRVALLKETMPQLADRPFSEARICWCA 349
: : : | :
Db 283 NYLLDYRLSGDKRLIYGGVVIYGARDPADIEAIRPMUKTFPQLKDVKIDFA---W-- 337
: : : | :
QY 350 DTANREFLIDRHQP-----YHSLVLCGCGARGFKYLPFSIGNLVTDAMEGK 395
: : : | :
Db 338 -TGNFLLTLRLPQVGRIGNDIYYQQ-----GCSCGHGVTYTHVAGKVLAEARLQG 386
: : : | :

RESULT 13
S16133
dimethylglycine dehydrogenase (EC 1.5.99.2) - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C:Accession: S16133
R.;Lang, H.; Polster, M.; Brandsch, R.
Eur. J. Biochem. 198, 793-799, 1991
A>Title: Rat liver dimethylglycine dehydrogenase. Flavinylolation of the enzyme in hepatocytes
A:Reference number: S16133; MUID:91266966; PMID:1710985
A:Accession: S16133
A:Molecule type: mRNA
A:Residues: 1-857 <LAN>
A:CROSS-references: UNIPROT.Q63342; EMBL:X55995; NID:X55995; PIDN:CAA39468.1; PID:g56688
C:Keywords: flavoprotein; oxidoreductase

Query Match 5.6%; Score 142.5; DB 2; Length 857;
Best Local Similarity 20.3%; Pred. No. 0.0065;
Matches 98; Conservative 74; Mismatches 160; Indels 151; Gaps 23;

QY 21 LIVGAGTGTSTALHLARRGYNTVLDPPVPSPSAISAGN-----DYNKY 65
: : : ||| ||| ||| :
Db 46 VIIIGGCCVGSLSAYHLKAGMRDVLLE---KSELTAGSTWHHAAGUTTVFHGINLUKKI 101
: : : ||| ||| ||| :
QY 66 ISSGOYSNNKDEIEVNEILAEEAFN--GWKNLDFPKPYHDTGLLMSACSQEGDLRLGV 123
: : : | :
Db 102 -----HYDSIKLYLERLEETQVVG-----FHQPSIRLATTPVERDEFKYQ 143
: : : | :
QY 124 V-RPEDPNLVELTRPPOFRKLAP-----EGVLQDGFPCKWGKFARGAGWAHARNALVAA 178
: : : | :
Db 144 MTRTNHWATEGYIIIEKIHLEHFLPLNMDDKILAGY-----NPQGDHIDPYSLTWAL 195
: : : | :

Qy 179 ARERQRMGV--KF---VTG-----TPQGRVVTLIFENNNDVKGAVTGDGKIWRPA 221
| : | : | : |||||
Db 196 ATGARKYGVLLKYPAPYTSKKRPDGTWIDVETPGQSV-----RA 234

Qy 222 ERIFLCAG--ASAGQFLDFKNQLRPTAWTLVHIALKPEERALYKNIPIVFNIE----- 272
| : | : | : |||||
Db 235 NRIVNAAGFWAREVGKRMIGLDHDLPLIPVOHQVTVTSTIPEVKALKRELPLVRDEGSYYLR 294

Qy 273 ---RGFFFEDEERGEIKICDE-----HPGYTNMVQSADGMTMSIPPEKTQIPKEAETR 323
| : | : | : |||||
Db 295 QERDGLLFGPVESQEKMKLQASWAHVAGVPFGFKELPESDLDR-----TEHVEAAEMM 348

Qy 324 VRALLKETMPOLADRPSPFARICWCADTANREFLIHRPHQYHSVLVGCASGRGFYLPS 383
| : | : | : |||||
Db 349 VPVLKADIINIINVNGPITYS-----PDILP---MVGPHQGVRNYWVAIG-FGYGIHAGG 399

Qy 384 IGNLIVD-AMEGKVPOKIHLELIKWNPDIAANNRNWDTLGRGGPNRVMDFHVDKWTNVQ 442
| : | : | : |||||
Db 400 VGKYSOWILHGEPD---FDLIELDPN-----RYG-----KWTTTQ 432

Qy 443 YRD 445
| :
Db 433 YTE 435

RESULT 14
B83078
C:Probable D-amino acid oxidase PA4548 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 16-Aug-2004
C:Accession: B83078
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizioquchi, S.D.; Warrenner, P.; Hickey, M.J.; Brj
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 939-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho:
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: B83078
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <STO>
A:Cross-references: UNIPROT:P33642; GB:AE004868; GB:AE004091; NID:g9950785; PIDN:AAG07939
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4548
C:Superfamily: Sarcosine oxidase

Query Match 5.5%; Score 139.5; DB 2; Length 364;
Best Local Similarity 24.6%; Pred. No. 0.0033;
Matches 103; Conservative 44; Mismatches 165; Indels. 107; Gaps 19;

Qy 17 SSSLLIVGAGTWCTSTALHLARRGYNTVLDPPVPSPAISAGDNVNKVISSGSXNNKD 76
| : : : ||||| | : : : |||||
Db 2 SERVVMVAGVIGILLTARETLAGRLVTLVERGESGREASWAGGIVSPLYPWRYSP--- 58

Qy 77 EIEVNEILABEAFNGWKNDPLPKPYHYDTGLIMLSACSQEGDLRLGVRRVPRGDDPNL---- 132
| : : : ||||| | : : : |||||
Db 59 -----AVTALAHSQD---FYD-----ALGQELLDETGL-----DPEVHTVG 92

Qy 133 ----VEL-----TRPEQRKLAPGCVLGQDFPGWKGYFARS--GAGWAHARNA 174
| : : : ||||| | : : : |||||
Db 93 LYWLDDDDQTEALQWARKHTRP---LKEVPITEEAAVPEGLGAGFORAYVMSCVANVRNP 149

Qy 175 LVAAAREAQRMGVKFTVTPQGRVVTLIFENNNDVKGAVTGDGKIWRAEITFLCAGASAQ 234
| : : : ||||| | : : : |||||
Db 150 RLARSRLASLOOFANLELHEQTVEGRWLDGRVVGVSATSRGEI-RGDKVLLAAGAWSGE 208

Qy 235 FLDFKNQLRPTAWTLVHIALKPEERALYK---NIPVFNIERGFFFEDEERGEIKICD 290
| : : : ||||| | : : : |||||
Db 209 L-----LKPLGLELPVPVVK-GQMILLYKCAADFIRPMV-LAKGRYAIIPRD-GHLLIGS 259

Qy 291 --EHFGYTNMVQSADGMTMSIPPEKTQIPKEAETRALLKETMPOLAD-----R 338

Db 260 TLEHSG-----FDKTP-TDEAQESLRASAAELLPDLAQMPVAHWAGLR 302
Qy 339 PFSPARICWADTANREFLDRHPQYHSLVLCGASGRGFKYLPDSIGNLIIVDAMEGKVP 397
Db 303 PGSPGEGIPY-----ICPVPGFDGLMINTGHRNGLVLADPASCRLLDLADMSGREP 351

RESULT 15
A83905
hypothetical protein BH2041 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 16-Aug-2004
C:Accession: A83905
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A83905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <STO>
A:Cross-references: UNIPROT:Q9KB87; GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB057
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2041
C:Superfamily: Sarcosine oxidase

Query Match 5.5%; Score 138; DB 2; Length 376;
Best Local Similarity 22.3%; Pred. No. 0.0046;
Matches 90; Conservative 62; Mismatches 165; Indels 66; Gaps 18;

Qy 20 LLIVAGTGTSTALHARGYTNVLDPPYVPSAISAGN-DVNKVISSGQYSNNKDEI 78
Db 7 LCIIIGGVIGLSIAHTAKKG-ANVLVVEKDRLGQEAAGGRAGILTTVAEG-----VDRH 60
Qy 79 EVNEILAEAFN---GWKNDPLFKPYHYDTGLLSACSQEGLDPLGVVRPG---EDPNL 132
Db 61 DPYSIFRSLSQVLLQW-----IPELEEKTIQISSGLVKONLLRLALHEQEAALLEDHWL 114
Qy 133 VELTRPQFRKLAP-----EGVLQGDPPGKGYFARSGAGWAHARNALV---AAAREAOR 184
Db 115 MQETNSQIKRLSAEEARTLEPLLSNDL-----IMALCSPDEYHLEPSLYLNALARWQ 169
Qy 185 MGKFTVTGTPQGRVVTLLIFENNVDVKGAVTGDGKIWRAERTFLCAGASA---GQFLDPKNQ 241
Db 170 LG---VTIREQTEVIQFIHQNNRMYGVTNTGTQIVAEITTLAAGAWSHLLSLRLAKVKVP 226
Qy 242 LRPTAWTLVHIALKPEREALYKNIPIVFNIERGFPFPEDEERGEIKICDEHPGYTNMVQS 301
Db 227 VYPLKGQL--IVLKQGD---TDLNVMINTSKGYILSKPNTGVVILGATSEKEKFNKFI-T 279
Qy 302 ADGTMMSIP-PEKTOIPKEAETVRALLKETMPQLADRPFSFARICWCADTANREFLDR 360
Db 280 AKGIQELLPLFD--YVPELKNWEIQLHWAGLRPA-----TPDQQPIIGY 321
Qy 361 HPQYHSLVLCGASGRGFKYLPDSIGNLIIVDAMEGKVPQKIHLEL 403
Db 322 VPWEGWYLATGHRHGVLLSGWTGHLVABEELEK--DQSHKL 362

Search completed: May 27, 2005, 13:48:47
Job time : 37.2303 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 13:38:36 ; Search time 55.3518 Seconds
(without alignments)
975.171 Million cell updates/sec

Title: US-10-622-893A-5
Perfect score: 2530
Sequence: 1 MGSSGDDDLALAVTKSSSL.....LEGCLPIPLNLTGHHHHH 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 538906 seqs, 114359116 residues

Total number of hits satisfying chosen parameters: 538906

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
 - 7: /cgn2_6/ptodata/2/paa/US11_NEW_COMB.pep.*
 - 8: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	226.5	9.0	459	8	US-60-643-717-13558
2	181.5	7.2	440	7	US-11-097-143-4818
3	177.5	7.0	390	6	US-10-990-328A-13084
4	177.5	7.0	445	6	US-10-990-328A-13083
5	169	6.7	387	6	US-10-990-477-6
6	169	6.7	389	6	US-10-534-583-1
7	165	6.5	387	6	US-10-990-477-2
8	150	5.9	774	6	US-10-990-328A-13788
9	150	5.9	866	6	US-10-990-328A-13787
10	145.5	5.8	388	6	US-10-526-324-372
11	130.5	5.2	386	6	US-10-526-324-29
12	126.5	5.0	862	1	PCT-US05-10680-20
13	126.5	5.0	862	7	US-11-096-051-20
14	125.5	5.0	570	6	US-10-800-077-386
15	124.5	4.9	530	8	US-60-669-241-28797
16	119	4.7	437	7	US-11-040-686-33
17	119	4.7	437	7	US-11-040-686-31
18	119	4.7	437	7	US-11-040-686-37
19	117.5	4.6	374	6	US-10-924-025A-118
20	117.5	4.6	459	7	US-11-040-686-24
21	117.5	4.6	459	7	US-11-040-686-26
22	117.5	4.6	459	7	US-11-040-686-31
23	115.5	4.6	389	1	PCT-IB03-04509-1503
24	115	4.5	460	6	US-10-996-104A-22
25	115	4.5	505	7	US-11-031-175-9986

26	112.5	4.4	407	6	US-10-996-104A-20	Sequence 20, Appl
27	111.5	4.4	879	6	US-10-204-639-58	Sequence 58, Appl
28	109	4.3	42	6	US-10-532-053-15	Sequence 15, Appl
29	108.5	4.3	320	1	PCT-US05-04621-10	Sequence 10, Appl
30	108.5	4.3	320	7	US-11-057-027-10	Sequence 10, Appl
31	108	4.3	796	6	US-10-450-763-48529	Sequence 48529, A
32	107.5	4.2	270	8	US-60-645-354-36	Sequence 36, Appl
33	107	4.2	421	7	US-11-040-686-42	Sequence 42, Appl
34	106.5	4.2	822	8	US-60-643-717-17883	Sequence 17883, A
35	105.5	4.2	296	1	PCT-US05-10680-18	Sequence 18, Appl
36	105.5	4.2	296	7	US-11-096-051-18	Sequence 18, Appl
37	105.5	4.2	822	8	US-60-643-717-7212	Sequence 7212, Ap
38	105	4.2	1664	6	US-10-450-763-44587	Sequence 44587, A
39	104.5	4.1	211	8	US-60-645-354-44	Sequence 44, Appl
40	104.5	4.1	339	7	US-11-122-943-10	Sequence 10, Appl
41	104.5	4.1	339	7	US-11-122-943-12	Sequence 12, Appl
42	104	4.1	273	8	US-60-645-354-40	Sequence 40, Appl
43	104	4.1	308	1	PCT-US05-02996-6	Sequence 6, Appl
44	104	4.1	308	7	US-11-046-644-6	Sequence 6, Appl
45	104	4.1	515	7	US-11-097-143-42765	Sequence 42765, A

ALIGNMENTS

RESULT 1

US-60-643-717-13558
; Sequence 13558, Application US/60643717
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53629)A
; CURRENT APPLICATION NUMBER: US/60/643,717
; CURRENT FILING DATE: 2005-01-12
; NUMBER OF SEQ ID NOS: 19247
; SEQ ID NO 13558
; LENGTH: 459
; TYPE: PRT
; ORGANISM: ASPERGILLUS NIDULANS FGSC A4
; US-60-643-717-13558

Query Match 9.0%; Score 226.5; DB 8; Length 459;
Best Local Similarity 19.8%; Pred. No. 1.2e-10;
Matches 90; Conservative 38; Mismatches 101; Indels 225; Gaps 11;

Qy	15	TKSSLLIVGAGTGTSTALHARRGYTNTVLD-PYPVPSAISAGDNVKNVSSGGYSN 73	Sequence 20, Appl
Db	5	SKSIPAIIGDGAFLSTALHLYQNGYTDITVLEQDEKIPPPYSAANYLNKIV----- 57	Sequence 10, Appl
Qy	74	NKDEIEVNEILABEAFNGWKNNDPLFKPYHDHTGLLMSACSOEGLDRLGVRVRPGEDNVLV 133	Sequence 133, Appl
Db	58	-REBY-----EDP--- 64	Sequence 64, Appl
Qy	134	ELTRPEQFRKLAPGVLGDPFGWKGYFARSAGAHARNALVAAREAQRMGVKFTGT 193	Sequence 193, Appl
Db	65	----- 64	Sequence 64, Appl
Qy	194	PQGRVVTLIENNDVKGAVTDGKIWAERTFLCAGASAGQFLDKQLRPTAWLVHIA 253	Sequence 253, Appl
Db	65	----- 64	Sequence 64, Appl
Qy	254	LKPEERALYKNI PVIFNIERGFFPEPDERGEIKICDEHPGYTMVMSADGTMMSIPFEK 313	Sequence 313, Appl
Db	65	-----LCPMGSGYIN-TDKATGVSHS-PFPV 95	Sequence 95, Appl
Qy	314	TQ-----IPKEATRVALLKETMPQADRPSPFAICWCADTANEFILDRP-QVHS 366	Sequence 366, Appl
Db	96	NQEASSGFLPAEDETRIRKLLQOTLPALARNPLVLKSLCWFADTKDSDFIIDFVPGSKGS 155	Sequence 155, Appl
Qy	367	LVLGCGASGRGFKYLPISIGNLIVDAMEGKVPQKTHELIKWNPDIANRNWRDITLGRFGGP 426	Sequence 426, Appl
Db	156	WVIESADSGHGFKNFPIVGSWTSLLSKHTPNSH---SGKPNLRANMGQSTYKNTGLP 212	Sequence 212, Appl

```
QY 427 NRVMDPHDVKEWTV-----QYRDISKLKGELEG 455
Db 213 WVLPSVQARRGFENKGLVHEYLPIRLKGLREG 246

RESULT 2
US-11-097-143-4818
; Sequence 4818, Application US/11097143
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4818
; LENGTH: 440
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-4818

Query Match 7.2%; Score 181.5; DB 7; Length 440;
Best Local Similarity 24.1%; Pred. No. 6.7e-07;
Matches 108; Conservative 57; Mismatches 149; Indels 135; Gaps 24;

QY 17 SSSILVAGTGTSTALHLARRGY-----TNTVLDPPYVPSAISAG-NDVNKVISSG- 69
Db 31 SCGLVIIGGGMGASSAFWLKSRALQLGRKINLVVER-----DAGYTSASTVLSVGG 83

QY 70 --QVSNKDEIEVNEILAEAFNGWKNDPLFKPYVYHDTGLLMSACSOEGL----- 117
Db 84 VRQOFSLAENIEMSLFGYNFVNGREHLGDVLDLCYQPNGLILA-SEKGAHILAKNSKLQ 142

QY 118 DRIGVRVRPGEDPNLVELTPEQFRK-----LAPEGVLQ-----DFPGWKGYFARSGAGWA 169
Db 143 NELGAR-----NELLGPEALRQRPFWLSTEGVELGCHGIDKEGWF-----PWA 186

QY 170 HARNALVAAREAQRMKVFTGTPQGRVVTLIFENND-----VKGAVTGDGKIWAERTF- 225
Db 187 -----LLMGYKKARALGANFANGSVWG-----FEMNDSGLSGAVVDGVDLQRTVKFD 236

QY 226 ---LCAGASAGQ---FLDFGNQLRPTAWTLVHIALKPEERALLY-----KNIPVI---FN 270
Db 237 TCVLGAYSGQVARLAGIDGKEAKESLSVALPVERKRYVVVSTQGNCRGLATPLT 296

QY 271 IE-RGFFPEPD-----ERGBIKICDEHPGTYTNMVQSADGTMSIPFEKTQ 315
Db 297 VDPDGTFRRDGLCGNFCGRSPNEDEPECETILDVDHG-----FE--- 338

QY 316 IPKEAETRVRAALKETWQADRPFSFARI-----CWCA-----DTANREFLIDRHPQVHSL 367
Db 339 -----TDVWPTLANRVAPESVKIOSSWAGFYDHNFTDANGVIGRHPHYSNL 385

QY 368 VLCCGASGRGFKYLPISIGNLIVD-AMEGK 395
Db 386 FTAAGFSGHGIQOTPAVGRAISGLIDGK 414

RESULT 3
US-10-990-328A-13084
; Sequence 13084, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328A
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13084
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-990-328A-13084

Query Match 7.0%; Score 177.5; DB 6; Length 390;
Best Local Similarity 21.7%; Pred. No. 1.2e-06;
Matches 86; Conservative 66; Mismatches 174; Indels 71; Gaps 16;

QY 21 LIVGAGTGTSTALHLARRGYTNVTLDPYVPSAISAGNDVNKVISSGQYSNNKDEIEV 80
Db 11 IVIGAGIQGCTAYHLAKH-RKRILLLEQFFLPHSRGSSHGQSRIIRKAYLE-----DF 63

QY 81 NEILAEAFNGWKNDPLFKPYVYHDTGLLMSACSOEGLDRLGVVRVPGEDPNLVELTRPEQ 140
Db 64 YTRMMECYQIWAQ-----LEHEACTQLH--RQTGLLLGMKENQELTKIQANLSRQRV 115

QY 141 FRKLAPEGVLQGDFF-----GMKGYFARSGAGWAHARNALVA---AARE---AORMGVK 188
Db 116 EHQLSSEELKQRFNIRLPRGVEGLDMSG-GVIYAYKALRALQDAIRQLGGIVRDEK 174

QY 189 FVTGTPQGRVVTLIFENNDVKGAVTGDGKIWAERTFLCAGASAGQFLDKQLRPTAWT 248
Db 175 VVEINP-GLLVT-----VKTTSRYQAKSLIVITAGPWTNQL-----LRPLGIE 216

QY 249 LVHIALKPEERALLYKNIPVIFNIERGEF-----FPEPDEERGEIKICDEHPG 294
Db 217 MPLQTLRINVCYWREVMVPGSYGVSAQFPCFLWLGLCFPHIYGLTGEYPLGMKVSYHHGN 276

QY 295 YTNMVQSADGTMSIPFEKTQIPKEAETRV-RALLKETWQADRPFSFARI-CWCA-DTAN 353
Db 277 H-----ADPEERDCPTARDI---GDVQILSSFVRDHLPLDKPEP-AVIESCMYTNTPD 326

QY 354 REFLIDRHPQVHSLVLCGASGRGFKYLPISIGNLIVD 390
Db 327 EQFLIDRHPKYDINIVIGAGFSGHGFKLAPVVGKILYE 363

RESULT 4
US-10-990-328A-13083
; Sequence 13083, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328A
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13083
; LENGTH: 445
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-10-990-328A-13083

Query Match      7.0%; Score 177.5; DB 6; Length 445;
Best Local Similarity 21.1%; Pred. No. 1.5e-06;
Matches 86; Conservative 66; Mismatches 174; Indels 71; Gaps 16;

QY 21 LVVAGTGTCTGTAHLARGYTNVTLDPVPSAISAGNDVNVKVISSGQYSNNKDEIEV 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 IVVAGTGTCTGTAHLARGYTNVTLDPVPSAISAGNDVNVKVISSGQYSNNKDEIEV 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 81 NEILAEAFNGKNDPLFKPYHYDTGLLMSACSGEGDLRLGVRVRPCGEDPNLVELTRPEQ 140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 YTRMHCEYQIAQ-----LEHAGTQLH--RTGILLGLMKENQELKTIQAWLSQRV 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 141 FRKLAPEGLVQDPP-----GWKGYFARGAGWAHARNALVA---AARE---AQRMGVK 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 EHQLSSEELKQRPNRLPRGEVGLDNGS-GVIYAKALRALQDAIRQLGGIVRDEK 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 189 FVTGTPQGRVVTLLIFENNVDKAVTGDKIWAERTFLCAGASAGQFLDPKQLRPTAWT 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 VWEINP-GLLV-----VKTTSRYQAKSLVITAGPWTNQL-----LRPLGIE 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 249 LVHIALKPEERALLYKNIPVFNIRGE-----FPEPDEERGIKICDEHPG 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 272 MPLQTLINVCYRWEMVPGSYGVSQAPCFPLWLGLCPHHIYGLPTGEYPLMKVSYHGN 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 295 YTNVQASAGTMSIPPEKTIQPKAEATRV-RALLKETMPQLADRPFSFARICWCADTAN 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 332 H-----ADPEERDCPTARDI---GDVQILSSFVRDHLPLDKPEP-AVIESCWYNTPTD 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 354 REFLLDRHPQYHSLVLCGASGRGFKYLPISGNLIVD 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 382 EQFLDRHPKYDNIVIGASGSGHGFKLAPVVGKILYE 418
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-10-990-477-6
; Sequence 6, Application US/10990477
; GENERAL INFORMATION:
; APPLICANT: KIKKOMAN CORPORATION
; APPLICANT: FURUKAWA, Keisuke
; APPLICANT: KAJIYAMA, Naoki
; TITLE OF INVENTION: MODIFIED SARCOSINE OXIDASES, MODIFIED SARCOSINE OXIDASE GENES,
; FILE OF INVENTION: AND METHODS FOR PREPARING THE MODIFIED SARCOSINE OXIDASES
; FILE REFERENCE: 261714USO
; CURRENT APPLICATION NUMBER: US/10/990.477
; CURRENT FILING DATE: 2004-11-18
; PRIOR APPLICATION NUMBER: JP 2003-387975
; PRIOR FILING DATE: 2003-11-18
; PRIOR APPLICATION NUMBER: JP 2004-184960
; PRIOR FILING DATE: 2004-06-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Bacillus sp. strain KS-11A
US-10-990-477-6

Query Match      6.7%; Score 169; DB 6; Length 387;
Best Local Similarity 21.4%; Pred. No. 6.4e-06;
Matches 89; Conservative 69; Mismatches 172; Indels 86; Gaps 20;

QY 20 LLVVGAGTGTCTGTAHLARGYTNVTLDPVPSAISAGNDVNVKVISSGQYSNNKDEIE 79
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 VVVGAGSMGMAAGYLAQGVKTLV-DADPPHTGSHHGDTIRIHA-YGEGREYVP 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 80 VNEILAEAFNGKNDPLFKPYHYDTGLL-----MSACSQEGL-----D 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 F-ALRAQELWYELENETHNK-IFTKTGVLVFGPKGESDFVAETMEAAEHSLTVLDLEGD 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 119 RLGRVVRPGEDPNLVELTRPEQFKLAPEGVLQDFFGKGYFARGAGWAHARNALVA 178
```

```
Db 123 EINTR-WPG-----ITVPENYNAI-----FEPNSGVLFSENCIRSY 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 179 AREAORMGVKFTGTGPOGRVVTLLIFENNVDKAVTGDKIWAERTFLCAGASAGQFLDF 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 158 RELAVAKGAKILTVT---RVDEFEVSQDVK-IQTANGS-YTADKLIVSMGAMSKLLSK 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 239 KNQLRPTAWTLVHALKP-----BERALYKN-----IPV-IPNIERGGFFEPDEERGE 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 LN-----LDIPLQYRQVVVGFDSNEAKYSNDVDYPAFMVEVPKGIYYGPFSGG- 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 IKICDEHPGTNNVQASAGTMSIPPEKTIQPKAEATRVALLKETMPQLADRPFSFARI 345
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 263 ---CGLKIGHTVQQLDPTINREFGAYO---EDESNLDRDFLEKYNPE-ANGELKRGAV 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 346 CWCADTANRREFLLDRHPQYHSLVLCGASGRGFKYLPISGNLIVD-AMEGKVPQKI 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 CMTKTPDEHFVDTHPEHSNVFVAAGFSGHGFKFSSWGEVLSQLATTGKTEHDI 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-10-534-583-1
; Sequence 1, Application US/10534583
; GENERAL INFORMATION:
; APPLICANT: KISHIMOTO, Takahide
; APPLICANT: SOGABE, Atsushi
; APPLICANT: OKA, Masanori
; TITLE OF INVENTION: MODIFIED SARCOSINE OXIDASE, PROCESS FOR PRODUCING THE SAME AND
; FILE OF INVENTION: REAGENT COMPOSITION USING THE SAME
; FILE REFERENCE: 235670
; CURRENT APPLICATION NUMBER: US/10/534.583
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: PCT/JP2003/014423
; PRIOR FILING DATE: 2003-11-13
; PRIOR APPLICATION NUMBER: JP2002-329427
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: JP2002-329428
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: JP2003-33641
; PRIOR FILING DATE: 2003-02-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Arthrobacter SP. TE1826
US-10-534-583-1

Query Match      6.7%; Score 169; DB 6; Length 389;
Best Local Similarity 21.1%; Pred. No. 6.4e-06;
Matches 91; Conservative 69; Mismatches 171; Indels 100; Gaps 20;

QY 12 LAVTKSSLLIVGAGTGTCTGTAHLARRGYTNVTLDPVPSAISAGNDVNVKVISSQY 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MSIKQDYDVIVVGAGSMGMAAGYVLSKQGVKTLV-DSFPHPTGSHHGDTIRIHA-Y 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 72 SNNKDEIEVNEILAEAFNGKNDPLFKPYHYDTGLL-----MSACSQEGL 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 GEGREYVVPF-ALRAQELWYELENETHNK-IFTKTGVLVFGPKGEAPFVAETMEAAEHSL 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 118 RLGRVVRPGEDPNLVELTRPEQFKLAPEGVLQDFFGKGYFARGAGWAHARNALVA 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 D-----VDLLEGSSEINKRWP-GVTVPE--NYNAIPEKN-SGVLFSENCIRA 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 178 AAREAORMGVKFTGTGPOGRVVTLLIPENNVDK-----GAVTGDGKIWAERTFLCAGASA 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 YRELAENGAKVLTYP---VEDFEIADFKIQTAYGSFT-----ASKLIYSMGAWN 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 233 GQFLDFKNQLRPTAWTLVHALKPEERALLYKNIPVFIENIERGGFFEPDEERGE----- 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 SKLLSKLN-----IEIPLQP-----YRQVV-----GFFECDEKYSNTHGYPA 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 286 --IKI-----CDEHPGTNNVQASAGTMSIPPEKTIQPKAEATRVALLKE 330
```

```
Db 247 FMVEVPTGIYGPSPGGCLKIGYHTYQCKIDPDTINREF---GIYPEDEGNIRKFLFET 303
Qy 331 TWPQLADRPFSFARICWCADTANREFLIDRHPQYHSLVLCGASGRGFKYLPDSIGNLIYD 390
Db 304 YMPG-ATGELKSGAVCMYTKTPDEHFVIDLHPQFSNVAAGFSHGFRFSSVVGTELQ 362
Qy 391 -AMEGKVPQKI 400
Db 363 LAVTGKTEHDI 373

RESULT 7
US-10-990-477-2
; Sequence 2, Application US/10990477
; GENERAL INFORMATION:
; APPLICANT: KIKKOWAN CORPORATION
; APPLICANT: FURUKAWA, Keisuke
; APPLICANT: KAJIYAMA, Naoki
; TITLE OF INVENTION: MODIFIED SARCOSINE OXIDASES, MODIFIED SARCOSINE OXIDASE GENES,
; TITLE OF INVENTION: AND METHODS FOR PREPARING THE MODIFIED SARCOSINE OXIDASES
; FILE REFERENCE: 261714US0
; CURRENT APPLICATION NUMBER: US/10/990,477
; CURRENT FILING DATE: 2004-11-18
; PRIOR APPLICATION NUMBER: JP 2003-387975
; PRIOR FILING DATE: 2003-11-18
; PRIOR APPLICATION NUMBER: JP 2004-184960
; PRIOR FILING DATE: 2004-06-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 2
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Construct
US-10-990-477-2

Query Match 6.5%; Score 165; DB 6; Length 387;
Best Local Similarity 21.4%; Pred. No. 1.4e-05;
Matches 89; Conservative 68; Mismatches 173; Indels 86; Gaps 20;

Qy 20 LLIVGAGTGTSTALHARRGYNVTVDLPYPVPSAISAGNDVKNVSSQYNNKDEIE 79
Db 7 VIVVGAGSMGWAAGYLLAKGVKTLV-DAFDPHTGSHGDTRIIRHA-YGEGREYVP 64
Qy 80 VNEILAEAFNGKNDPLFPYHDTGLL-----MSACSQBGL-----D 118
Db 65 F-ALRAQELWYLENETHNK-IFTKTGLVFGPKGESDFVAETMEAAAEHSLIVDLLEG 122
Qy 119 RLGVVRPQGEDPNLVELTRPEQFKLAPEGVLOQDPGKGYFARSGAGWAHARNALVAA 178
Db 123 EINTR-WPG-----ITVPENYNAI-----FEPNSGVLFSENCIRSY 157
Qy 179 AREAQRMGKVFVTGTPGCRVVTLIFENNVDKAVGTGDKIWRARTPLCAGASAGQFLDF 238
Db 158 RELAVAKGAKLITYT---RVDEPVSQDVK-IQTANGS-YTADKLIVSGAWNSKLISK 212
Qy 239 KNQLRPTAWTLVHIALKP-----EERALYKN-----IPV-IFNIERGFFFPDEBERGE 285
Db 213 LN-----LDIPLQYQYVGVGFFDSNEAKYSNDVDYPAFMVVEPKGIYYGFPSPFG- 262
Qy 286 IKLCDEHPGYTNVQSGADGTMMSIPPEKTOIPKEAETRVALLKETPQLADRPFSFARI 345
Db 263 ---CGLKIGYHTYQCKIDPDTINREFGAYQ---EDESNLRDFLFKYMPPE-ANGELKKGAA 315
Qy 346 CWCADTANREFLIDRHPQYHSLVLCGASGRGFKYLPDSIGNLIYD-AMEGKVPQKI 400
Db 316 CMYTKTPDEHFVIDTHEHSNVFVAAGFSHGFRFSSVVGVELSQLATTGKTEHDI 371

RESULT 8
US-10-990-328A-13788
```

```
; Sequence 13788, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328A
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13788
; LENGTH: 774
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-990-328A-13788

Query Match 5.9%; Score 150; DB 6; Length 774;
Best Local Similarity 20.8%; Pred. No. 0.00066;
Matches 102; Conservative 77; Mismatches 187; Indels 124; Gaps 23;

Qy 2 GSGGDDDDLALAVTK---SSSLIVGAGTGTSTALHARRGYNVTVDLPYPVPSAISA 58
Db 31 GREGEKPPLSAETQWKDRAETVILGGCGVGSVLAYHLAKAGKDVLLVLE---KSELTA 86
Qy 59 GN-----DYNKVISSQYNNKDEIEVNEILAEAFN--GWKNDPLFKPY 101
Db 87 GSTWHAAGLTTFPHPGINLKKI-----HYDSIKLYEKLSEETQGVVG----- 128
Qy 102 YHDTGLLMSACSOEGLDRLGVRV-RPGEDPNLVELTRPEQFKLAP-----EGVLOQDFPG 156
Db 129 FHQPGSIRLATTTPVRVDEFKYQWTRTGMHATEQYLLIEPEKIOEMFFLLNNKVLAGLY-- 186
Qy 157 WKGYPARSAGWAHARNALVAAAREAQRMGVKFTVGTGQGVVTLIFENNVDKAVGTG 216
Db 187 -----NPGDGHIDPYSLTWALAAGARKCGALLKYAP---VTSLKASDGTWDVETPQ 237
Qy 217 KIWAERTFLCAG---ASAGQFLDFKNQLRPTAWTLVHIALKPEERALYKNIPVFNIE- 272
Db 238 SM-RANRIVNAAGFWAREVGKMGIGLEHPLIPVQHQQVVTSTIPEVKALKRELPLVRLDLE 296
Qy 273 -----RGFPFPEPPEERGEIKICDE-----HPCYTNMV--QSADGTMMSIPFEKTOI 316
Db 297 SYYLROERDGLLFGPYESQEKMKVQDSWVTNGVPPGFKELFESDLDRIMEHI----- 349
Qy 317 PKEAETRVALLKETPQLADRPFSFARICWCADTANREFLIDRHPQYHSLVLCGASGR 376
Db 350 -KAAMENVPVKKADIINVNGPITYS-----PDILP---MVGPHQGVRYNYWAIG-FGY 399
Qy 377 GFKYLPDSIGNLIYD-AMEGKVPQKIHELKWNPDIAANRNWRDTLGRFGGPNRMDFHDV 435
Db 400 GIHAGGVGKGLSDWILLHGEP--FDLIELDPN-----RYG----- 433
Qy 436 KEWNTVOYRD 445
Db 434 -KWTQTQYTE 442

RESULT 9
US-10-990-328A-13787
; Sequence 13787, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328A
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13787
; LENGTH: 866
```



```
; PRIOR APPLICATION NUMBER: US To be assigned
; PRIOR FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: 60/557,978
; PRIOR FILING DATE: 2004-03-30
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 20
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US05-10680-20

Query Match      5.0%; Score 126.5; DB 1; Length 862;
Best Local Similarity 21.3%; Pred. No. 0.073;
Matches 71; Conservative 34; Mismatches 103; Indels 125; Gaps 15;

Qy 196 GRVVTLIFENNVDKGVGTGDKIWRARTFLCAGASAGQFLDFKNQLRPTAWTLVHIALK 255
Db 599 GASLTLVFPERSPP---LTQYHTVWIPWNVF-----YVMDTLVMK 634
Qy 256 PEER-----ALYKNIPVIFNIERGEFFPEPDEERGEIKICDEHPGYTNMVQSADGTM 307
Db 635 KEENDIPSCDLSGFPVRNPPIVSSPLSTFFRSPSPEDSPI----- 673
Qy 308 SIP-----FEKTQIP-----KEAETRVALLKETMPQLADRPSPFARI----- 345
Db 674 -IPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQ-SIIPFNLMKVLHMLVA 731
Qy 346 ----CWCADTAN--REFLIDRHPQYHSLVLCGAS--GRGPKYLPISGNLIVDAMEGKVP 397
Db 732 RLFOKWFPASPENLAYTFIWDKTDAYNOKYVGLSEAVSVGYEY----- 774
Qy 398 QKTHELIKWNPDIAANRNWRDTLGRFGCPNRMVDHDKVKEWTVNQYRDISKLKE----- 452
Db 775 ESCLDLTWEKRTAILQGYELDASNMGG--WTLDKHHV---LDVQNGILYKNGENQFIS 829
Qy 453 -----LEGLPINPLL-----RTGHHHHH 472
Db 830 QPPVWSLEGKPIPNPLLGLDSTRTGHHHHH 862

RESULT 13
US-11-096-051-20
; Sequence 20, Application US/11096051
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacLachlan, Timothy K
; APPLICANT: Rastelli, Luca
; APPLICANT: Vernet, Corine
; APPLICANT: Eitenberg, Seth
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use
; FILE REFERENCE: Attorney Docket No. Cura 967
; CURRENT APPLICATION NUMBER: US/11/096,051
; CURRENT FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: 10/038,854
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 10/455,772
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: 60/557,978
; PRIOR FILING DATE: 2004-03-30
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 20
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-096-051-20

Query Match      5.0%; Score 126.5; DB 7; Length 862;
Best Local Similarity 21.3%; Pred. No. 0.073;
Matches 71; Conservative 34; Mismatches 103; Indels 125; Gaps 15;

Qy 196 GRVVTLIFENNVDKGVGTGDKIWRARTFLCAGASAGQFLDFKNQLRPTAWTLVHIALK 255

; PRIOR APPLICATION NUMBER: US/10800077
; GENERAL INFORMATION:
; APPLICANT: Reddy, Ramachandra
; APPLICANT: Gill, Parkash
; TITLE OF INVENTION: NUCLEIC ACID COMPOUNDS FOR INHIBITING
; FILE REFERENCE: VASG-P01-001
; CURRENT APPLICATION NUMBER: US/10/800,077
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: US 60/454,432
; PRIOR FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: US 60/454,300
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 396
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 386
; LENGTH: 570
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Recombinant B4ECv3 protein
US-10-800-077-386

Query Match      5.0%; Score 125.5; DB 6; Length 570;
Best Local Similarity 23.0%; Pred. No. 0.05;
Matches 49; Conservative 22; Mismatches 63; Indels 79; Gaps 11;

Qy 295 YTNMVQSADGTM-----SIPFETQIPKEAETRVALLKETMPQLAD-----RPFSAR 344
Db 402 YTFEVTALGVSSLATGVPFEPVNVTTD-----REVPPAVSDIRVTRSPSSUS- 451
Qy 345 ICWCADTANREFLIDRHPQYHSLVLCGASG-----RGPKYLPIS 384
Db 452 LANAVPRAPSGAWLDYEVKYEK-----GAGPSPSVFLKTSNPRAELRGLKRGASYLVQV 507
Qy 385 GNLIIVDAMEGKVPQKTHELIKWNPDIAANRNWRDTLGRFGCPNRMVDHDKVKEWTVNQYR 444
Db 508 -RARSEAGYGFFGQEHHS---QTQLDSESGWREQ---GSKRAIL----- 544
Qy 445 DISKLKGELEGLPIPNPLL-----RTGHHHHH 472
Db 545 -----QIEGKPIPNPLLGLDSTRTGHHHHH 570

RESULT 15
US-60-669-241-28797
; Sequence 28797, Application US/60669241
```

GENERAL INFORMATION:
APPLICANT: Monsanto Technology LLC
APPLICANT: Baum, James A
APPLICANT: Gilbertson, Larry A
APPLICANT: Kovalic, David K
APPLICANT: LaRosa, Thomas J
APPLICANT: Lu, Maolong
APPLICANT: Munyikwa, Tichifa R. I.
APPLICANT: Roberts, James K
APPLICANT: Wu, Wei
APPLICANT: Zhang, Bei
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CONTROL OF INSECT INFESTATION IN PLANTS
FILE REFERENCE: 38-21(53596)
CURRENT APPLICATION NUMBER: US/60/669,241
CURRENT FILING DATE: 2005-04-07
PRIOR APPLICATION NUMBER: 60560842
PRIOR FILING DATE: 2004-04-09
PRIOR APPLICATION NUMBER: 60565632
PRIOR FILING DATE: 2004-04-27
PRIOR APPLICATION NUMBER: 60579062
PRIOR FILING DATE: 2004-06-11
PRIOR APPLICATION NUMBER: 60603421
PRIOR FILING DATE: 2004-08-20
PRIOR APPLICATION NUMBER: 60617261
PRIOR FILING DATE: 2004-10-11
NUMBER OF SEQ ID NOS: 50011
SEQ ID NO 28797
LENGTH: 530
TYPE: PRT
ORGANISM: Lygus hesperus
FEATURE:
NAME/KEY: misc feature
LOCATION: (519)..(519)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
OTHER INFORMATION: Coding regions on vcdna: vcdna=SeqID_3881; Strand=+; Position=84-
FEATURE:
OTHER INFORMATION: Homolog annotation: Hit ID=XP_319829.1; Match level="QueryCoverag
OTHER INFORMATION: =97%, HitCoverage=96%, E-value=1e-136, Identity=45%"; Hit descrip
OTHER INFORMATION: ENSANGP00000015949 [Anopheles gambiae]
FEATURE:
OTHER INFORMATION: Pfam annotation: Pfam_ID=DAO; Match level="Score=211.2, E-value=2
OTHER INFORMATION: -60, Copies=1"; pfam description=FAD dependent oxidoreductase
US-60-669-241-28797

Query Match 4.8%; Score 124.5; DB 8; Length 530;
Best Local Similarity 21.4%; Pred. No. 0.055;
Matches 94; Conservative 64; Mismatches 167; Indels 115; Gaps 24;

QY 20 LLIVGAGTGTGTALHLARRGYTNVTVL----DPYVPVPSAISAGNDVVK-----VI 66
DB 108 ILIIGGWMGFSIAYWLQQRALDHLRIVVVEQDPTVNP-----DKSKWLEYQKSTSVL 160
QY 67 SSG---QYSNNKDEIEVNEILAEAFN-----GWKN-----DPLPKPYHDTGLLMSACSQ 114
DB 161 SVGGLRQQFSLKENIELSMASAFMRNKHHLGVGEGADPPDICFNP----CGYLFLEA-SE 215
QY 115 EGL----DRLGVRVRGEGEDNLVELTPEQFRKLAP----EGVLQGDFFPGWKGYFARSGA 166
DB 216 RGTHLQDNVNLQTSIGAK---VELLSPEKLKPKFPMTNDGVA-----LGCHGLENE 265
QY 167 GWAAHARNALVAAAREAQRMGVKFTVTPQGRVVTLIFENN-----D 207
DB 266 GWDFPWSLLSAFRRKAVALETVV----NGKVTSFDFMKNKSPIYNOYFAPGEYSHVEKAN 321
QY 208 VKGAVTGDGKIWAERTFLCAGASAGO---FLDFKNQLRPTAWTLVHIAKPEERALLY-K 263
DB 322 VK-LEDGSVKSIXFSQLIIVAAGCNTGEVGLKMGSGAGPLG---APIEPRKRYVFMV 377
QY 264 NIPVIFNIERGFEPD-----BERGEIKIDEHPGYTNVQSDGTWMSIPPEKTQIP 317
DB 378 HCPDAPGLAMPFIIDPTGAVYRRGEGVGLGSLGSGDEBQBPCCDN--MDVDYEPFE-- 433

QY 318 KEAETRVRALLKETMPQLADR--PFSFARI--CWCA-----DTANREFLIDRHPOYHSLVL 369
DB 434 -----NKLWPIIANRVPAFECAKVKNAWSGFYDFNSFDENAFIGPHPAFHNHV 482
QY 370 GCGASGRGFKYLPISIGNLIV 389
DB 483 VAGFSGHGIQQSPAVGRAMM 502

Search completed: May 27, 2005, 13:51:58
Job time : 57.3518 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2005, 21:41:14 ; Search time 921 Seconds
(without alignments)
9459.321 Million cell updates/sec

Title: US-10-622-893A-6

Perfect score: 1419

Sequence: 1 atgggaggttcgggtgacga.....atcacccatcatcattaa 1419

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5700845 seqs, 3069779757 residues

Total number of hits satisfying chosen parameters: 11401690

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1419	100.0	1419	19	US-10-622-893A-6
2	335.2	23.6	1320	19	US-10-470-678-1
3	333.6	23.5	1320	19	US-10-470-678-3
4	332	23.4	1320	19	US-10-470-678-2
5	332	23.4	1320	19	US-10-470-678-4
6	62.8	4.4	1314	16	US-10-232-655-4
7	57	4.0	422	18	US-10-767-701-22052
8	50.4	3.6	1314	16	US-10-232-655-2
9	38.8	2.7	3731	17	US-10-600-230-78
10	37.2	2.6	1302	9	US-09-815-242-9647
11	37	2.6	951	17	US-10-282-122A-25526
					Sequence 6, Appli
					Sequence 1, Appli
					Sequence 3, Appli
					Sequence 2, Appli
					Sequence 4, Appli
					Sequence 4, Appli
					Sequence 2, Appli
					Sequence 78, Appli
					Sequence 22052, A
					Sequence 2, Appli
					Sequence 78, Appli
					Sequence 9647, Ap
					Sequence 25526, A

12	36.6	2.6	660	14	US-10-123-155-480	Sequence 480, App
13	36.6	2.6	660	15	US-10-146-731-480	Sequence 480, App
14	36.6	2.6	660	15	US-10-140-472-480	Sequence 480, App
15	36.6	2.6	660	15	US-10-141-761-480	Sequence 480, App
16	36.6	2.6	660	16	US-10-142-885-480	Sequence 480, App
17	36.6	2.6	660	16	US-10-158-790-480	Sequence 480, App
18	36.6	2.6	660	17	US-10-137-871-480	Sequence 480, App
19	36.6	2.6	660	17	US-10-140-923-480	Sequence 480, App
20	36.6	2.6	660	17	US-10-141-756-480	Sequence 480, App
21	36.6	2.6	660	17	US-10-141-759-480	Sequence 480, App
22	36.6	2.6	660	17	US-10-140-805-480	Sequence 480, App
23	36.6	2.6	660	17	US-10-140-864-480	Sequence 480, App
24	36.6	2.6	660	17	US-10-142-426-480	Sequence 480, App
25	36.6	2.6	671	14	US-10-184-644-345	Sequence 346, App
26	36.2	2.6	671	14	US-10-184-634-346	Sequence 346, App
27	36.2	2.6	2634	18	US-10-437-963-18493	Sequence 18493, A
28	36	2.5	2113	18	US-10-425-115-182534	Sequence 182534, A
29	35.8	2.5	739	14	US-10-123-155-478	Sequence 478, App
30	35.8	2.5	739	15	US-10-146-731-478	Sequence 478, App
31	35.8	2.5	739	15	US-10-140-472-478	Sequence 478, App
32	35.8	2.5	739	15	US-10-141-761-478	Sequence 478, App
33	35.8	2.5	739	16	US-10-142-885-478	Sequence 478, App
34	35.8	2.5	739	16	US-10-158-790-478	Sequence 478, App
35	35.8	2.5	739	17	US-10-137-871-478	Sequence 478, App
36	35.8	2.5	739	17	US-10-140-923-478	Sequence 478, App
37	35.8	2.5	739	17	US-10-141-756-478	Sequence 478, App
38	35.8	2.5	739	17	US-10-141-759-478	Sequence 478, App
39	35.8	2.5	739	17	US-10-140-805-478	Sequence 478, App
40	35.8	2.5	739	17	US-10-140-864-478	Sequence 478, App
41	35.8	2.5	739	17	US-10-142-426-478	Sequence 478, App
42	35.6	2.5	414	10	US-09-918-995-33483	Sequence 33483, A
43	35.4	2.5	3659	18	US-10-437-963-2456	Sequence 2456, Ap
44	35.2	2.5	882	14	US-10-184-644-574	Sequence 574, App
45	35.2	2.5	882	14	US-10-184-634-574	Sequence 574, App

ALIGNMENTS

RESULT 1

US-10-622-893A-6
; Sequence 6, Application US/10622893A
; Publication No. US20050014935A1
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Sheng
; APPLICANT: Datta, Abhijit
; APPLICANT: Wang, Yuping
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DETERMINATION OF GLYCATED PROTEINS
; FILE REFERENCE: 46692001300
; CURRENT APPLICATION NUMBER: US/10/622,893A
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1419
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding a chimeric protein
US-10-622-893A-6

Query Match	100.0%;	Score 1419;	DB 19;	Length 1419;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1419;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATCGGAGGTTCCGGTGACGATGATGACCTCGGCTCGCCGCTCACTAGTCAATCATCTCTC	60	
Db	1	ATCGGAGGTTCCGGTGACGATGATGACCTCGGCTCGCCGCTCACTAGTCAATCATCTCTC	60	
Qy	61	CTGATCGTTGGTCCCGGACTTGGGGCACTTCAACGGCTCTGCACCTCGCGCGCGGGA	120	
Db	61	CTGATCGTTGGTCCCGGACTTGGGGCACTTCAACGGCTCTGCACCTCGCGCGCGGGA	120	

```
QY 121 TATACCAAGCTTACCGTGTGGACCCCTATPCTGTCTCCCTAGCCCATCTCCGCGGAAAC 180
Db 121 TATACCAAGCTTACCGTGTGGACCCCTATPCTGTCTCCCTAGCCCATCTCCGCGGAAAC 180
QY 181 GAGGTGAACAAAGTCATTAGCAGTGGCCAAATATTCGAATAACAAAGACGAAATCGAAGTG 240
Db 181 GAGGTGAACAAAGTCATTAGCAGTGGCCAAATATTCGAATAACAAAGACGAAATCGAAGTG 240
QY 241 AATGAGATCTTGGCGGAAGAGCGTTTAAAGGTTTGAAGAACACCCGCTTTTCAACCG 300
Db 241 AATGAGATCTTGGCGGAAGAGCGTTTAAAGGTTTGAAGAACACCCGCTTTTCAACCG 300
QY 301 TATTATCATGATACGGGCCCTGCTGATGCTGTCTGCTCGCAGAGGCGCTTGATCGCCCTG 360
Db 301 TATTATCATGATACGGGCCCTGCTGATGCTGTCTGCTCGCAGAGGCGCTTGATCGCCCTG 360
QY 361 GCGGTCCGGGTACGTCGGCGGAGGATCTTAATCTCGTGGAACCTTACCCGCGCGGAGCAA 420
Db 361 GCGGTCCGGGTACGTCGGCGGAGGATCTTAATCTCGTGGAACCTTACCCGCGCGGAGCAA 420
QY 421 TTTCTGTAATCGCCCGGAGCGGTGTCGAAGGTGATTTTCCGGTTCGAAAGGGTAC 480
Db 421 TTTCTGTAATCGCCCGGAGCGGTGTCGAAGGTGATTTTCCGGTTCGAAAGGGTAC 480
QY 481 TTTGCGCGTTCGCGCGCTGGCTGGGCACATGCAAGGAATGCCCTTAGTGGCAGCAGCACGC 540
Db 481 TTTGCGCGTTCGCGCGCTGGCTGGGCACATGCAAGGAATGCCCTTAGTGGCAGCAGCACGC 540
QY 541 GAAGCACAGCGCATGGGTGTAATTTGTTATCTGGCACCCCGCAGGGTCTGTAGTCACG 600
Db 541 GAAGCACAGCGCATGGGTGTAATTTGTTATCTGGCACCCCGCAGGGTCTGTAGTCACG 600
QY 601 TTAATCTTTGAAAATACAGATGTAAAGGTGCGTTACGGCGCATGCGCAAAATTTGCGA 660
Db 601 TTAATCTTTGAAAATACAGATGTAAAGGTGCGTTACGGCGCATGCGCAAAATTTGCGA 660
QY 661 GCGGAACGTACATTCCTGTGTGCTGGGGCTAGCGCGGTGAGTTCCTAGATTTCAAGAAT 720
Db 661 GCGGAACGTACATTCCTGTGTGCTGGGGCTAGCGCGGTGAGTTCCTAGATTTCAAGAAT 720
QY 721 CAATCTCGACCAACCGCTTGGACCCCTGGTACACATTCGCGTTAAACCGGAAGAACGTGCG 780
Db 721 CAATCTCGACCAACCGCTTGGACCCCTGGTACACATTCGCGTTAAACCGGAAGAACGTGCG 780
QY 781 TTGTACAAAATATACGGTTATCTTTAATCATCGAACGGGGTTCCTTTGTAACCCGAT 840
Db 781 TTGTACAAAATATACGGTTATCTTTAATCATCGAACGGGGTTCCTTTGTAACCCGAT 840
QY 841 GAGGAGCGCGGTGAGATTAAATATGCGATGAACACCCCGGCTACACAAATATGGTCCAG 900
Db 841 GAGGAGCGCGGTGAGATTAAATATGCGATGAACACCCCGGCTACACAAATATGGTCCAG 900
QY 901 AGTGCGACGCGCAGATGAGATTCGTTTCGAAAAAAACCCAGATTCGAAAGAACGC 960
Db 901 AGTGCGACGCGCAGATGAGATTCGTTTCGAAAAAAACCCAGATTCGAAAGAACGC 960
QY 961 GAAAACGGCGTTCCGGCCCTGCTGAAGAGACAAATGCCCCAGCTGGCAGACCCCTCCATTC 1020
Db 961 GAAAACGGCGTTCCGGCCCTGCTGAAGAGACAAATGCCCCAGCTGGCAGACCCCTCCATTC 1020
QY 1021 AGCTTGCACGCAATTTGCTGGTGCGGATACCGCGAATCGCGAATTCCTGTAGATCGA 1080
Db 1021 AGCTTGCACGCAATTTGCTGGTGCGGATACCGCGAATTCCTGTAGATCGA 1080
QY 1081 CATCCGAGTACCAAGTCTTGTGTTGGGCTGTGGTCCGAGCGGAAGAGGGTTTAAATAT 1140
Db 1081 CATCCGAGTACCAAGTCTTGTGTTGGGCTGTGGTCCGAGCGGAAGAGGGTTTAAATAT 1140
QY 1141 CTGCCTCTATTGGGAATCTCATGTTGACCGCATGGAAGGTAAAGTCCGCGCAAAATTT 1200
Db 1141 CTGCCTCTATTGGGAATCTCATGTTGACCGCATGGAAGGTAAAGTCCGCGCAAAATTT 1200
```

```
QY 1201 CACGAATTAATCAAGTGGAAACCCGACATTCGGCGGAACCGTAACCTGGCGTGATCTCTG 1260
Db 1201 CACGAATTAATCAAGTGGAAACCCGACATTCGGCGGAACCGTAACCTGGCGTGATCTCTG 1260
QY 1261 GGGCGTTTGGCGGTCCAAATCGTGTGATGGAATTTTCATGATGTGAAGNAATGGACCAAT 1320
Db 1261 GGGCGTTTGGCGGTCCAAATCGTGTGATGGAATTTTCATGATGTGAAGNAATGGACCAAT 1320
QY 1321 GTTCAGTATCGTGTATATTTTCCAAAGCTGAAAGGAGAGTTCGAAAGGTAAAGCAATCCCTAAC 1380
Db 1321 GTTCAGTATCGTGTATATTTTCCAAAGCTGAAAGGAGAGTTCGAAAGGTAAAGCAATCCCTAAC 1380
QY 1381 CGGTACTCGCACAGGCCATCACCATCATCATTA 1419
Db 1381 CGGTACTCGCACAGGCCATCACCATCATCATTA 1419

RESULT 2
US-10-470-678-1
; Sequence 1, Application US/10470678
; Publication No. US20050101771A1
; GENERAL INFORMATION:
; APPLICANT: KOUZUMA, Takuji et al.
; TITLE OF INVENTION: COMPOSITION FOR ASSAYING GLYCATED PROTEINS
; FILE REFERENCE: 1516-0121P
; CURRENT APPLICATION NUMBER: US/10/470,678
; PRIOR FILING DATE: 2003-07-30
; PRIOR APPLICATION NUMBER: JP 2001/ 22953
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: JP 2001/ 39796
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: JP 2001/ 24002
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Fusarium oxysporum IFO-9722
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1320)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: DDBJ E16562
; DATABASE ENTRY DATE: 1999-07-28
; PATENT DOCUMENT NUMBER: JP 1998201473-A/1
; PATENT FILING DATE: 1997-01-20
; PUBLICATION DATE: 1998-08-04
; US-10-470-678-1

Query Match 23.6%; Score 335.2; DB 19; Length 1320;
Best Local Similarity 55.3%; Pred. No. 3.4e-99;
Matches 721; Conservative 0; Mismatches 568; Indels 15; Gaps 3;

QY 41 TCATCAAGTCATCATCTCTCTGATCGTTGGTGGCGGACCTTGGGGACCTCAACGGCTC 100
Db 11 TCACCAACACAGTCCCAAAATTCATCGTTGGTGGCGGAACTTGGGGATGCTCAACTGCC 70
QY 101 TGCACCTCGCGCGCCGGGATATACCAAGTTACCGTGTCTGGACCCCTATCTCTGTCCTTA 160
Db 71 TCCATCTCGCCGCTCGGGGTACACCAACGTCATCTGTTCTCGATGTCAATTCGATCCCGT 130
QY 161 GCGCCATCTCCGCGGAAACGAGCTGAACAAAGTCATTAGCAGTGGCAATATTCGAATA 220
Db 131 CACCGATATCAGCGGGGCATGATGTAAACAACTTCTGCTGGCCGACATGTCGACTGCCGATA 190
QY 221 ACAAAA-----GACGAAATCGAAGTGAATGAGATCTTGGCGGAAGAGCGGTTTAAACGGTT 274
Db 191 GCAAAAGTGATGATGAAGACTCAATCTGGAAGAAGCACTTAGCTACGCCGAGCTCAAGGAT 250
QY 275 GGAAGAACACCGCTTTTCAACCGGTATATCATGATACGGGCGCTGCTGATGCTCTGCTT 334
Db 251 GGCTCCACGACCCCTGTCTTCCAAACCATCTTGCCACAATACAGGCTCTGTCTGGCTGGCT 310
```



```
QY 635 TTACGGCGCATGCAAAATTTGGAGAGCGGAACGTATCTCTGTGTGCTGGGCTAGCG 694
Db      |||||
QY 611 AGACGGCAGATGTTAGGAGCACAGAGCGGATCGAACTATCTTCGCTGGTCTTCAG 670
Db      |||||
QY 695 CGGGTCAGTTCCTAGATTTCAAGAACTTCGACCAACCGCTTCGACCCCTGGTACACA 754
Db      |||||
QY 671 CAGAGTTCCTCTCGATTTTGAGAACCCAGATCCAGGCTAGCGGCTGGACCCCTGGGCCATA 730
Db      |||||
QY 755 TTCCGTTTAAACCGGAAGACGTTGTGTACAAAATATACCGGTTATCTTTAAACATCG 814
Db      |||||
QY 731 TCAGATGACACAGAGAAACCAAGCTGTACAGAACCTGCCACCTCTTTTCAACATCA 790
Db      |||||
QY 815 AACGGGGTTTTCTTTTGAACCGGATGAGAGCGCGTGAGATTTAAATATGCGATGAAC 874
Db      |||||
QY 791 ACCAAGTTTTCTTCATGGAACTGATGAGGATCTTCATCAACTCAAGATGTGCGATGAAC 850
Db      |||||
QY 875 ACCGGGCTACACAAATATGTTCCAGATGTCAGACGGCAGC-----ATGATGAGCATTCGGT 931
Db      |||||
QY 851 ATCCGGGCTACTGCAACTGGGTTGAAAAGCCTGTTCTTAAGTACCCCACTCCATCCCT 910
Db      |||||
QY 932 TCGAAAAAACCCAGATTTCCAAAAGAGCGGAACGGCGTTCCGGGCCCTGCTGAAAAGAGA 991
Db      |||||
QY 911 TCGAAAGCATCAAGTGCCAAACGAGGCTGAACGACATGAAGCAGTTTTCTGAAAAGATA 970
Db      |||||
QY 992 CAATGCCCCAGCTGGCAGACCGTCCATTTCAAGCTTTCCGACGCAATTTGCTGTGTCGCGATA 1051
Db      |||||
QY 971 TCATGCTCAGCTTGACAGATCGGCGCTTGTTCATGCTCGAATCTGCTGTGTCGCTGATA 1030
Db      |||||
QY 1052 CCGCGAATCCGGAATTCCTGATAGATCGACATCCGAGTAGACACAGCTTCTGTGTGTTGGGCT 1111
Db      |||||
QY 1031 CACAGGATAGAAATGTTCTCTGATCACCTATCATCTCGACATCCCTCACITTTGTCATTTGCTT 1090
Db      |||||
QY 1112 GTGCTGCGACGGAAGAGGGGTTAAATATCTGCTTCTATTTGGGAATCTCATTTGTGACG 1171
Db      |||||
QY 1091 CAGGTGATTCGGGACCGGTTACATGCAATATCATCAATCAATTTGAAAAGTTTCATCTGACT 1150
Db      |||||
QY 1172 CGATGGAAGTAAAGTGGCCGCAAAAATTCACGAATTTAATCAAAGTGGAAACCCGACATTG 1231
Db      |||||
QY 1151 GTATGAGGGTACGCTTGAGGAAGGTTTGCCAAAGTTCTCGAGATGGCGACAGAGAAT 1210
Db      |||||
QY 1232 CCGGGAACCGTAATGCGCTGATATCTCTGGGGGTTTGGCGGTTCMAATCGTGTGATGG 1291
Db      |||||
QY 1211 TTACCGAGTTCTGGGGTAAAGATCTCTCGATCGGTTTGGAGCTGACGATAAGATCATGG 1270
Db      |||||
QY 1292 ATTTTC-----ATGATGTCAAGGAATGGAACCAATGTTCAAGTAT 1329
Db      |||||
QY 1271 ATTTGCCAAGATGATGTAGAGGGATGGACAAATATCAAGAAT 1314
Db      |||||
```

RESULT 4

```
US-10-470-678-2
; Sequence 2, Application US/10470678
; Publication No. US20050101771A1
; GENERAL INFORMATION:
; APPLICANT: KOJUZIMA, Takuji et al.
; TITLE OF INVENTION: COMPOSITION FOR ASSAYING GLYCATED PROTEINS
; FILE REFERENCE: 1516-0121P
; CURRENT APPLICATION NUMBER: US/10/470,678
; PRIOR FILING DATE: 2003-07-30
; PRIOR APPLICATION NUMBER: JP 2001/ 22953
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: JP 2001/ 39796
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: JP 2001/ 24002
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence derived from Fusarium oxysporum IFO-9722
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1320)
;
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (1114)..(1115)
;
US-10-470-678-2

Query Match      23.4%; Score 332; DB 19; Length 1320;
Best Local Similarity 55.1%; Pred. No. 3.9e-98;
Matches 719; Conservative 0; Mismatches 570; Indels 15; Gaps 3;

QY 41 TCACCTAAGTCATCATCTCTCTGATGTTGGTGGCGGACCTTTGGGACACCTCAACCGCTC 100
Db      |||||
QY 101 TGAACCTCGCGGCGCGGATATACCAAGTTTACCGTGTGACACCCCTATCTGTCTCCCTA 160
Db      |||||
QY 71 TCCATCTCGCCGCTCGGGGTTACCAACAGTCACTGTTCTCGATGTCAATTCGCATCCCGT 130
Db      |||||
QY 161 GCGCCATCTCCGCGGAAACGAGTGAACAAAGTCAATTAGCAGTGGCCCAATATTTCGAATA 220
Db      |||||
QY 131 CACCGATATCAGCCGGGCGATGATGTAAACAAACTTCTGCGCGACATGTGCGTCCGATA 190
Db      |||||
QY 221 ACAA-----GACGAAATCGAAGTGAATGAGATCTTGGCGGAAGAGGCTTTAAACGGTT 274
Db      |||||
QY 191 GCAAAGGTGATGATGAAGACTCAATCTGGAAGCACTTAGTACGCCGCGAGCTCAAGGAT 250
Db      |||||
QY 275 GGAAGAACGACCGCTTTTCAAACCGTATATCATGATAGCGGCCCTGTGATGTCTGCTT 334
Db      |||||
QY 251 GGTCTCACGACCTCTGTCTTCCAACCAATCTTGCACAAATACAGGCTCTGCTGGCTGGCT 310
Db      |||||
QY 335 GCTCGCAGAGAGGCGCTGGATCGCTGGGCGTCCGGGTACGTCCGGCGCAGGATCCCTAATC 394
Db      |||||
QY 311 CAAACCAAGTCTATCAAGCAGCTGTGTAGAGATGAGATCGGTGACGATCAGCACAGT 370
Db      |||||
QY 395 TGTGTGAATTTACCCGCGCGAGCAATTTCTGAAATCGCCCCCGGAAGCGGTGTTTCAAG 454
Db      |||||
QY 371 ATACACCTCTCAACACAGCAGAGAAGATTTTCAGAAAAGACCATGCTGAGGGTATCTCTGACAG 430
Db      |||||
QY 455 GTGATTTTCCGGGTTGGAAGGGTACTTTGCGGGTTTCCGGCGTCTGGGCGACATGCCAA 514
Db      |||||
QY 431 GTAACCTTTCAGGCTGGAAGGGCTTTTACAAGCCCAACGGGTTCTGGTTCGGTTCATGCTC 490
Db      |||||
QY 515 GGAATGSCCTTAGTGGCAGCAGCAGCGAGCAAGCAGCGCATGGGTGTAAATTTGTTACTG 574
Db      |||||
QY 491 GAAAAGCTATGAAGCTGCTTTTCGAAGAGAGCGAGAGGCTTGGTGTCAAAATTCATCACTG 550
Db      |||||
QY 575 GCACCCCGCAGGGTCTGTAGTCACGTTAATCTTTGAAAATTAACGATGTAAAGAGTGCCG 634
Db      |||||
QY 551 GCTCTCCCGAAGGAAGGTGGAGAGTCTGATCTTTGAAGACGGCGATGTTTCGAGGTGCCA 610
Db      |||||
QY 635 TTACGGCGCATGCAAAAATTTTGAGAGCGGAACGATCATTTCTGTGTCTGGGCTAGCG 694
Db      |||||
QY 611 AGACGGCAGATGTTAAGGAGCACAGAGCGGATCGAACTATCTTCGCTGGTCTTCAG 670
Db      |||||
QY 695 CGGGTCAGTTCCTAGATTTTCAAGAACTTCGACCAACCGCTTCGACCCCTGGTACACA 754
Db      |||||
QY 671 CAGAGTTCCTCTCGATTTTGAAACCAAGATCCAGCTAGCGGCTAGACCCCTGGGCGATA 730
Db      |||||
QY 755 TTGCGTTAAAAACCGGAAGAACGTTGTGTACAAAATATACCGGTTATCTTTTAAACATCG 814
Db      |||||
QY 731 TCCAGATGACACCGAAGAAACCAAGCTGTACAAAGAACCTGCCACCTCTTTTCAACATCA 790
Db      |||||
QY 815 AACGGGGTTTTCTTTTGAACCGGATGAGAGCGCGGTGAGATTTAAATATGCGATGAAC 874
Db      |||||
QY 791 ACCAAGGTTTCTTTCATGGAACTGATGAGGATCTTCATCAACTCAAGATGTGCGATGAAC 850
Db      |||||
QY 875 ACCGGGCTACACAAATATGTTCCAGATGTCAGACGGCAGC-----ATGATGAGCATTCGGT 931
Db      |||||
QY 851 ATCCGGGCTACTGCAACTGGGTTGAAAAGCCTGTTCTTAAGTACCCCACTCCATCCCT 910
Db      |||||
QY 932 TCGAAAAAACCCAGATTTCCAAAAGAGCGGAACGGCGTTCCGGGCCCTGCTGAAAAGAGA 991
Db      |||||
```

Db 911 TCAGAAAGCATCAAGTCCACCGAGGCTGAAACGACGATGAAGCAGTTTCTGAAAGATA 970
Qy 992 CAATGCCAGCTGGCAGACCGTCCATTCAGCTTCGACGCGATTGCTGGTGTCGCCGATA 1051
Db 971 TCATGCCCTCAGCTTCGACATCGGCGCGTGTTCATGCTCGAATCTGCTGGTGCGCTGATA 1030
Qy 1052 CCGCGAATCGGAATTCCTGTATAGATCGACATCCGACGATACCACAGTCTTGTGTGGGCT 1111
Db 1031 CACAGGATAGAATGTTCTGTATACCTATCATCTCGACATCCCTCACTGTCTATTGCTT 1090
Qy 1112 GTGTGCGAGCGGAAGAGGGTTTAAATATATGCTCTTCTATTGGGAATCTCATTTGTCAGC 1171
Db 1091 CAGGTGATTGGGCACCGGTTACTGGCAATATCATCATCAATGGAAAGTTTCACTCTGACT 1150
Qy 1172 CGATGGAAGTAAAGTCCCGCAAAATTCACGAAATTAATCAAGTGGAAACCCGACATG 1231
Db 1151 GTATGGAGGGTACGCTTGAAGGAAGGTTTGCAAGTCTTGAGATGGCGACGAGAAAT 1210
Qy 1232 CGGCGAACCGTAACTGCGGTGATACCTCTGGGCGGTTTGGCGGTCGAAATCGTGTGATGG 1291
Db 1211 TTACCGAGTTCTGGGGTAAAGATCCTCTGGATCGGTTTGGAGCTGACGATAAGATCATGG 1270
Qy 1292 ATTTTC-----ATGATGTGAAGGAATGGACCAATGTTTCAGTAT 1329
Db 1271 ATTTGCCAAGATGATGATAGAGGATGGACAAATATCAAGAAT 1314

RESULT 5

US-10-470-678-4
; Sequence 4, Application US/10470678
; Publication No. US20050101771A1
; GENERAL INFORMATION:
; APPLICANT: KOZUUMA, Takuji et al.
; TITLE OF INVENTION: COMPOSITION FOR ASSAYING GLYCATED PROTEINS
; FILE REFERENCE: 1516-0121P
; CURRENT APPLICATION NUMBER: US/10/470,678
; CURRENT FILING DATE: 2003-07-30
; PRIOR FILING DATE: 2001-07-30
; PRIOR FILING DATE: 2001-01-13
; PRIOR FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence derived from Fusarium oxysporum IFO-9722
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (1114)..(1115)

US-10-470-678-4

Query Match 23.4%; Score 332; DB 19; Length 1320;
Best Local Similarity 55.1%; Pred. No. 3.9e-98;
Matches 719; Conservative 0; Mismatches 570; Indels 15; Gaps 3;
Qy 41 TCATAAGTCATCATCTCTCTGTATCGTTGGTGGCGGACCTGGGGACCTCAACGGCTC 100
Db 11 TCACCAACAGTCCCAAAATTCATCGTTGGTGGCGGAATCTGGGGATGCTCAACTGCC 70
Qy 101 TGCACTTCGGCGCGCGGATATACCAACGTTACCGTGGGACCCCTATCTCTGCTCCTA 160
Db 71 TCCATCTCGCGCGTGGGGTTTACACCAACGTCACCTGTTCTCGATGTCAATCGCATCCGT 130
Qy 161 GCGCCATCTCCGCGGAAACGACGTGAACAAAGTCATTAGCAGTGGCGCAATATTCGAATA 220

Db 131 CACCGATATCAGCGCGGCATGATGTAAACAAACTTGTCTGCGCGACTGTGCACTGCCGATA 190
Qy 221 ACAA-----GACGAATCGAAGTGAATGAGATCTTTGGCGGAAGAGCGCTTTAACGGTT 274
Db 191 GCAAGGTGATGATGAAGACTCAATCTGGAAGACACTTAGCTAGCGCCGACGCTCAAGGAT 250
Qy 275 GGAAGAACGACCCGCTTTTCAAACCGTATTATCATGATACGGGCGCTGCTGATGTCTGCTT 334
Db 251 GGTCCACGACCCCTGTCTTCCAAACCAATCTGCGCCAAATACAGGCTCTGTCTGGCTGGCT 310
Qy 335 GCTCGAGAGGGGCTGCGATCGCTGGGCGTCCGGGTACGTCGCGGCGAGGATCCTTAATC 394
Db 311 CAACACCAAGTCTATCAACGAGCTGGTAGAAGATGAGATCGGTGAGCAGCATCGACCAGT 370
Qy 395 TGGTGAACATTACCCGCGGAGCAATTTTCGTAACCTGGCCCCGGGAAGGCGTGTGCAAG 454
Db 371 ATACACCTCTCAACACAGCAGAGAATTTTCAGAAAGACCATGCTCGAGGGTATCTCGACAG 430
Qy 455 GTGATTTTCCGGGTTGGAAGGGTACTTTTCGGCGTTCGGGCGTCCGGGCTGGCGCAGCATGCAA 514
Db 431 GTAACTTTCCAGGCTGGAAGGGCTTTTACAAGCCCAACGCGGTTCTGGTTGGGTTTCATGCTC 490
Qy 515 GGAATGCTTAGTGGCAGCAGCAGCGAAGCAGCATGGGTGTAATAATTTGTTACTG 574
Db 491 GAAAGCTATGAAGCTGCTTTTGAAGAGAGCAGAGGCTTGGTGTCAAAATTCATCACTG 550
Qy 575 GCACCCCGCAGGGTCTGTAGTCACTTAATCTTTGAAAAATAACGATGTAAGAGGTGCCG 634
Db 551 GCTCTCCGGAAGGAAGGTGGAGAGTCTGATCTTTGAAGACGCGGATGTTTCGAGGTGCCA 610
Qy 635 TTAGCGCGGATGGCAAAATTTGGAGAGCGGAAAGTACATTCCTGTGTGCTGGGCTAGCG 694
Db 611 AGACGCGAGATGGTAAAGGAGCAGACAGAGCGGATCGAACTATTTTTCGCGCTGGCTTCAG 670
Qy 695 CGGTTCAGTTCTCTAGATTTTCAAGAACTCACTTCGACCAACCGCTTGGACCTGTGTACACA 754
Db 671 CAGAGTTCTTCTCGATTTTGAGAACCGATTCAGCTCAGCGTGGCGTGGACCTGGGCCATA 730
Qy 755 TTGCGTTTAAACCGGAAGAACGTCGCTTGTGTACAAAATAATATACCGGTTTATCTTTAAACATCG 814
Db 731 TCCAGATGACACCAAGAAACCAAGCTGTACAAAGAACCTGCCACCTCTTTTCAACATCA 790
Qy 815 AACGGGGGTTTTTCTTTGAACCGGATGAGAGCGGCTGAGATTAAATATGCGATGAAC 874
Db 791 ACCAAGGTTTTCTTCATGGAAACCTGTATGAGGATCTTCACTCAACTCAAGATGCGCATGAAC 850
Qy 875 ACCGGGCTACACAAATATGTTCCAGATGTCAGACGCGACG---ATGATGAGCATTCGCT 931
Db 851 ATCCGGGCTACTGCAACTGGGTTGAAAAGCTGTGTTCTAAGTACCCCGACGTTCCATCCCT 910
Qy 932 TCGAAAAAACCCAGATTTCCAAAAAGAACCGCAACCGCGCTTCGGGCGCTGCTGAAAAGAGA 991
Db 911 TCGAAGACATCAAGTGCCCAACCGAGGCTGAAACGACGATGAAGCAGTTTCTGAAAGATA 970
Qy 992 CAATGCCCCAGCTGGCAGACCGTCCATTACGTTTCGACGCGATTTGCTGGTGTCGCGATA 1051
Db 971 TCATGCTCAGCTTGCAGATCGGCGCGCTTGTTCATGCTCGAAATCTGCTGGTGCGCTGATA 1030
Qy 1052 CCGCGAATCGGAATTTCTGTATAGATCGACATCGGAGTACCACAGTCTTGTGTGGGCT 1111
Db 1031 CACAGGATAGAATGTTCTGCTGATCACCTATCATCTCGACATCCCTCACTTGTCTATTGCTT 1090
Qy 1112 GTGTGCGAGCGGAAGAGGGTTTAAATATATGCTCTTCTATTGGGAATCTCATTTGTCAGC 1171
Db 1091 CAGGTGATTGGGCAACGGGTTACGTGCATATCATCAATTTGNAAGTTTCACTCTGACT 1150
Qy 1172 CGATGGAAGTAAAGTGGCGCAAAAAATTCAGAAATTAATCAAGTGGAAACCCGACATG 1231
Db 1151 GTATGGAGGGTACGCTTGAAGGAAGGTTTGCCAAAGTTCTGGAGATGGCGACGAGAAAT 1210
Qy 1232 CGCGGAACCGTAACTGCGGTGATACCTCTGGGCGGTTTGGCGGTCGAAATCGTGTGATGG 1291


```
; FILE REFERENCE: 227590US0
; CURRENT APPLICATION NUMBER: US/10/232,655
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Coniochaeta sp.
US-10-232-655-2

Query Match      3.6%; Score 50.4; DB 16; Length 1314;
Best Local Similarity 52.3%; Pred. No. 4e-05;
Matches 136; Conservative 0; Mismatches 121; Indels 3; Gaps 1;

Qy 628 GTGTCCTTACGGCGGATGCGAAATTTGGAGCGGAAAGTACATTCCTGTGTGCTGGG 687
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
604 GCGCTTGAGACGGCAGATGGTACCAATATTACCTGACAAAGGTGCTTAGCAGCTGGC 663
Qy 688 GCTAGCGGGGTGAGTCTCTAGATTTCAGAAATCACTTCGACCAACGGCTTGACCCCTG 747
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
664 GCATGGAGCCCAACCCCTGGTGGACCTCGGAATCAATGTTGCTCGAAGGCTTGGGTGTAT 723
Qy 748 GTACACATTCGTTTAAACCGGAAGACGTCGCTGTGTACAAAATATACCGGTTATCTTT 807
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
724 GCTCATATTCAATGACGCTCGAGAGCGGCTGAGTATAGGTGTCCAGTTGTGTAT 783
Qy 808 AACATCAACGGGGTTTTCTTTGAACCCGATGAGGAGCGCGTGAGATTAATAATGC 867
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
784 AATGGCGAATTTGGCTTCTTTTGAGCCTAAT--GAGTTTGTGTGAATAAAGGTGTGC 840
Qy 868 GATGAACACCGGGCTACAC 887
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
841 GACGAATTCAGGATTTCTC 860

RESULT 9
US-10-600-230-78/c
; Sequence 78, Application US/10600230
; Publication No. US20040092020A1
; GENERAL INFORMATION:
; APPLICANT: Wilkinson, Jack
; APPLICANT: McBride, Kevin
; APPLICANT: Bertain, Sean
; TITLE OF INVENTION: GENETIC CONSTRUCTS HAVING HETEROLOGOUS
; TITLE OF INVENTION: 3' POLYADENYLATION SIGNAL SEQUENCE MOTIFS THAT FUNCTION IN
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 0325.210
; CURRENT APPLICATION NUMBER: US/10/600,230
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: 60/390,529
; PRIOR FILING DATE: 2002-06-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 3731
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-600-230-78

Query Match      2.7%; Score 38.8; DB 17; Length 3731;
Best Local Similarity 46.6%; Pred. No. 0.48;
Matches 124; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

Qy 1061 GCGAATTCCTGATAGATCGACATCCGACGATACACAGTCTTGTGTGGGCTGTGGTSCGA 1120
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3297 GCGACTTCATATACCAACCAATCCATCCGACCCACACCTCGTCTCTTGCAACTGGTGCA 3238
```

```
Qy 1121 GCGAAGAGGGTTTAAATATCTGCTTCTATTTGGGAATCTCATTTGTTGACCGGATGGAAG 1180
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3237 GCGGCCACGCATACAAATTTCTCCCACTTATCGGCGAAAAAGGTGTCGATCGCTGGAGG 3178
Qy 1181 GTAAAGTGCCGCAAAAAAATTCAGCAATTAATCAAGTGGAAACCCGGACATTTGCGGCGAACC 1240
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3177 GTAAAGTCCGACCCGAACTGCGGCGCTCTGGGACTGCGCGAGGCTGTGGACAGAGGCC 3118
Qy 1241 GTAACTGGCGTGATCTCTCGGGCGTTTTTGGCGGTCCTCAAAATCGTGTGATGGAATTTTCATG 1300
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3117 ACTTTGACGGAACAGAGGATGGAAGTCGTTTCAGGGGATAAGCGGCTGCTGCTTTTGGATG 3058
Qy 1301 ATGTGAAGGAATGACCAATGTTTCAG 1326
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3057 AGTTGGCGAAGGGGCGAGAAAGCGAAG 3032

RESULT 10
US-09-815-242-9647
; Sequence 9647, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9647
; LENGTH: 1302
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1302)
US-09-815-242-9647

Query Match      2.6%; Score 37.2; DB 9; Length 1302;
Best Local Similarity 52.6%; Pred. No. 0.92;
Matches 81; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
```

```
Qy 31 GCCTCGCGCTCACTAAGTCATCATCTCTCTGATGTTGTCGCGGACTTGGGGCACC 90
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 GCTCTGCAAGACCATATTGATCGGATGCGGTGATCATTTGCGGTGCTTTTCAGGCATC 126
Qy 91 TCAACGGCTCTGCACCTCGCGCGCGCGATATACCAACGTTACCGTGTGTCGACCCCTAT 150
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
127 AATACCGCGCTGGAATCTGCGGAGCAGGGGATCACTAATGTTGTCGTTCTTGGAAAGCCGC 186
```



```
Db 279 CSCKDPTPIEFSPDPLPDKNLVNPVAVIAGNRPNLYRLRLSLLSAQGVSPQMITY 335
QY 473 AAGGCTACTTTTGGCGCTTCGGCGCTGGTGGGCACATGCAAGGAATCCCTTAGTGCAG 532
Db 336 FIDGYEPEPMDVVALFGLRGIQHTPIISIKNARVSQHYKASLTATFNLFPFAVLEED 395
QY 533 CAGCAGCGAAGCAGCAGCATCGGTGTAAATTTGTTACTGGCACCCCGCAGGTCGTG 592
Db 396 LDIADVDFSLQSIIHLLEDDSLYCISAWNDQGYEHTAEDPALLYRVETMPLGLWVLR 455
QY 593 TAGTCACGTTAATCTTTGAAAATAACGATGTAAGGTGCGCTTACGGCGCATGGCAAAA 652
Db 456 SLYKEELEPKWPTPEKLDWDMWMMRMEQRRGECIIPDVSRSYHFGIVGLNNGYFHEA 515
QY 653 TTTGGAGAGCGGAGCATCTCTGTGCTGGGCTAGCGGCTAGCTTCTTAGATT 712
Db 516 YFKKHENTVPGVQLRNVDSLKEAYEVEVHRLLSAEVLDHKNPCEDSFLPDTEGHTY 575
QY 713 TCAAGAAATCAACTTCGACCAACCGCTTGGACCCCTGGTACACATTTGCGTTAAAAACCGGAAG 772
Db 576 VAFIRMEKDDFTTTWTOLAKCLHLWDLVGRGNHRLWRLFRKKNHFLVGVGPASPSYVK 635
QY 773 AACGTGCGTTGTACAAAAATA 793
Db 636 PPSVTPFIFFLEPPPKKEGAPCA 656
```

RESULT 15

```
US-10-141-761-480
; Sequence 480, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 480
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-761-480
```

```
Query Match 2.6%; Score 36.6; DB 15; Length 660;
Best Local Similarity 8.5%; Pred.No. 1;
Matches 53; Conservative 193; Mismatches 372; Indels 3; Gaps 1;

QY 173 CCGAAACGACGTCGACAAAGTCATTAGCAGTGCCCAATATTCGAATAACAAAGACGAAA 232
Db 39 CQTGAVLFLVTVIVNKLILDTERRAISEANEDPEPEQDYDEALGRLEPPRRRSGPRV 98
QY 233 TCGAAGTGAATGATCTTTCGGCGAAGAGCGGTTTACGGTTGGAAGAACGACCCGCTTT 292
```

```
Db 99 LDVEVYSSKVVAVDGTTVLLEDEAREQGRGHIHVILNQATGHVMAKRVDFDYSPEHE 158
QY 293 TCAAAACCGTATTATCATGATACGGGCTGCTGATGTCTGCTCGCAGAGGGCTGG 352
Db 159 AMVLFNMPVAPGRVLICTVKDESGFLKDTAKALLRSLSQAGPALGWRDWTWAFVGRKG 218
QY 353 ATGCGCTGGCGCTTCGGGTACGTCGGGCGGAGATCCTAATCTGTGGTGAATTTACCGGCC 412
Db 219 PVFGEKHSKSPALSSMGDPVLLKTDVPLSSABEAECHEWADTELNNRRRRFCSKVEGYGV 278
QY 413 CGGAGCAATTCGTAAACTGGCCCCGGAAGCGGTGTGCAAGGTGATTTCCGGGTGGA 472
Db 279 CSCKDPTPIEFSPDPLPDKNLVNPVAVIAGNRPNLYRLRLSLLSAQGVSPQMITY 335
QY 473 AAGGCTACTTTTGGCGCTTCGGCGCTGGTGGGCACATGCAAGGAATCCCTTAGTGCAG 532
Db 336 FIDGYEPEPMDVVALFGLRGIQHTPIISIKNARVSQHYKASLTATFNLFPFAVLEED 395
QY 533 CAGCAGCGAAGCAGCAGCATGGGTGTAAATTTGTTACTGGCACCCCGCAGGTCGTG 592
Db 396 LDIADVDFSLQSIIHLLEDDSLYCISAWNDQGYEHTAEDPALLYRVETMPLGLWVLR 455
QY 593 TAGTCACGTTAATCTTTGAAAATAACGATGTAAGGTGCGCTTACGGCGCATGGCAAAA 652
Db 456 SLYKEELEPKWPTPEKLDWDMWMMRMEQRRGECIIPDVSRSYHFGIVGLNNGYFHEA 515
QY 653 TTTGGAGAGCGGAAACGATACATCTCTGTGCTGGGCTAGCGGCTAGCTTCTTAGATT 712
Db 516 YFKKHENTVPGVQLRNVDSLKEAYEVEVHRLLSAEVLDHKNPCEDSFLPDTEGHTY 575
QY 713 TCAAGAAATCAACTTCGACCAACCGCTTGGACCCCTGGTACACATTTGCGTTAAAAACCGGAAG 772
Db 576 VAFIRMEKDDFTTTWTOLAKCLHLWDLVGRGNHRLWRLFRKKNHFLVGVGPASPSYVK 635
QY 773 AACGTGCGTTGTACAAAAATA 793
Db 636 PPSVTPFIFFLEPPPKKEGAPCA 656
```

Search completed: May 30, 2005, 01:42:31

Job time : 923 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2005, 19:56:49 ; Search time 267 Seconds
(without alignments)
8696.167 Million cell updates/sec

Title: US-10-622-893A-6
Perfect score: 1419
Sequence: 1 atggggaggttcgggtgacga.....atcaccatcatcatcatataa 1419

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685	48.3	1314	2	US-09-031-059-2
2	37.8	2.7	7218	1	US-08-232-463-14
3	35.8	2.5	390416	4	US-09-949-016-16923
4	34.4	2.4	1296	4	US-09-902-540-5414
5	34.4	2.4	34953	4	US-09-902-540-1263
6	33.8	2.4	468	4	US-09-252-991A-7672
7	33.8	2.4	1032	4	US-09-252-991A-7887
8	33.8	2.4	2739	4	US-09-252-991A-7592
9	33.8	2.4	2778	4	US-09-252-991A-7958
10	33.8	2.4	147382	4	US-09-949-016-14624
11	33.6	2.4	281	4	US-09-313-294A-4144
12	33.6	2.4	4403765	3	US-09-103-840A-2
13	33.6	2.4	4411529	3	US-09-103-840A-1
14	33.4	2.4	801	4	US-09-252-991A-9143
15	33.4	2.4	1818	4	US-09-252-991A-8941
16	33.2	2.3	4181	1	US-07-670-611-1
17	33.2	2.3	4181	1	US-08-220-674-1
18	33.2	2.3	4181	1	US-08-445-186-1
19	33.2	2.3	4181	1	US-08-446-549-1
20	33.2	2.3	4181	2	US-08-446-550-1
21	33	2.3	1431	4	US-09-489-039A-5719
22	33	2.3	3931	4	US-10-101-464A-961
23	33	2.3	3939	4	US-10-101-464A-961
24	32.8	2.3	999	4	US-09-902-540-8104
25	32.8	2.3	6890	4	US-09-902-540-828
26	32.8	2.3	4403765	3	US-09-103-840A-2
27	32.8	2.3	4411529	3	US-09-103-840A-1

C 28	32.6	2.3	2230	4	US-09-620-312D-282	Sequence 282, App
C 29	32.6	2.3	60424	4	US-09-949-016-12175	Sequence 12175, A
C 30	32.4	2.3	328	4	US-09-313-294A-3938	Sequence 3938, A
C 31	32.4	2.3	3108	4	US-09-614-221A-332	Sequence 332, App
C 32	32.2	2.3	774	4	US-09-252-991A-9809	Sequence 9809, App
C 33	32.2	2.3	1200	4	US-09-252-991A-8724	Sequence 8724, App
C 34	32.2	2.3	1347	4	US-09-489-039A-4659	Sequence 4659, App
C 35	32.2	2.3	174639	4	US-09-949-016-16509	Sequence 16509, A
C 36	32	2.3	1346	4	US-09-107-532A-261	Sequence 261, App
C 37	31.8	2.2	364	4	US-09-621-976-17202	Sequence 17202, A
C 38	31.8	2.2	1047	4	US-09-902-540-4084	Sequence 4084, App
C 39	31.8	2.2	24602	4	US-09-902-540-1202	Sequence 1202, App
C 40	31.8	2.2	35574	4	US-09-949-016-11843	Sequence 11843, A
C 41	31.8	2.2	35574	4	US-09-949-016-14511	Sequence 14511, A
C 42	31.6	2.2	1187	4	US-10-000-489-105	Sequence 105, App
C 43	31.6	2.2	1205	4	US-09-976-594-457	Sequence 457, App
C 44	31.6	2.2	129899	4	US-09-949-016-14684	Sequence 14684, A
C 45	31.6	2.2	148156	4	US-09-949-016-11776	Sequence 11776, A

ALIGNMENTS

RESULT 1
US-09-031-059-2
; Sequence 2, Application US/09031059
; Patent No. 5948659
; GENERAL INFORMATION:
; APPLICANT: KATO, NOBUO
; APPLICANT: SAKAI, YASUYOSHI
; APPLICANT: TANI, YOSHIKI
; APPLICANT: FUKUYA, HIROSHI
; TITLE OF INVENTION: RECOMBINANT FRUCTOSYL AMINO ACID
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,059
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/899,172
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, MARC S
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 0020-4253P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1311
US-09-031-059-2

Query Match 48.3%; Score 685; DB 2; Length 1314;

Best Local Similarity 70.7%; Pred. No. 9.9e-230;
Matches 926; Conservative 0; Mismatches 380; Indels 3; Gaps 1;

QY 38 CCGTCACTAAGTCATCATCTCTCTGATCGTGGTCCGGGACCTTGGGGCACCTCAACGG 97
DB 5 CAGTCACCAAGTCTTGGTGCATATGATCATCGGGCGGGACCTGGGGTGGCTCACTG 64
QY 98 CTCTGACCTCGCGCGCGGGATATACAAAGTTACCGTGTGAGACCCCTATCTGTGCC 157
DB 65 CCCTGCACTTGGCCCGCAGAGGATACCAAAATGTCACTGTCTTTCGACCCGTACCCGGTTC 124
QY 158 CTAGCGCCATCTCCGCGGAAACGAGCTGAAACAAAGTCATTAGCAGTGGCCAAATTTCGA 217
DB 125 CATACGCCATTTCCGCGCGGCAACGAGCTCAACAAAGATCATCTCGTCCGGCGAGTACAGCA 184
QY 218 ATAAACGAACGAAATCGAAGTGAATGAGATCTTGGCGGAAGAGCGGTTTAAACGGTTGGA 277
DB 185 GCAGAAGACGAGGTGCGAAGTCAATGAGATTATCGCCGAACAGGCCCTTCAATGGCTGGA 244
QY 278 AGAACGACCGCTTTTCAACCGGTATTATCATGATACGGGCGCTGCTGATGTCTGTGCT 337
DB 245 AAAATGACCCCATCTTTCAAGCGGTACTACCAACGACCGGCGTCTGTGATGTTCGGCCACCA 304
QY 338 CGCAGGAGGCTTGGATCGCTGGGCGTCCGGGTAGCTCCGGCGCAGGATCTTATCTGG 397
DB 305 CACAGGAAGATTGGAGCGCTCTGGGGGTCCGCGTCCGACCTGAAGATGAACCCGATGTAG 364
QY 398 TGAACATTACCCGCCCGGAGCAATTTCTGTAATCTGCTGCGGCGGAGCGGTTTGAAGGTG 457
DB 365 CCGAATTGACTCGGCGGAGCAGTTCGCCACGCTGGCCCC---GGCGTCTTGAAGGTA 421
QY 458 ATTTTCGGGTGGAAAGGGTACTTTTTCGCGCTTCCGGCGCTGGCTGGGCACATGCAAGGA 517
DB 422 ACTTCCCGGTTGGAGGGGTACCAATTCGCTCAAAACGCGGCTGGCGCATGCGGCA 481
QY 518 ATGCTTTAGTGGCAGACGACGCGAAGCAGCGCATGGTGTATAAATTTGTTACTGGCA 577
DB 482 ACGCCCTTGGTTCGCGCGGCGGAGGACAGCGCTGGGTGTGGCGCTTCGTCGCGGGAT 541
QY 578 CCGCGCAGGGTCTGTAGTCACTGTTAATCTTTGAAATAAACGATGATAAAGGTGCGGTTA 637
DB 542 CGCGCAGGCGAGATCATCGTTGATTTTGAAGAACACGATGTAAGGGTGCCTGCA 601
QY 638 CGGCGCATGCGCAAAATTTGGAGAGCGGAACGTACATTCCTGTGTGCTGGGGCTAGCGGG 697
DB 602 CGCGGACGCAAGATCTGCGGCGCGAGCAGACTATCTCTGCGCTGTGTGCGCGCGCG 661
QY 698 GTCACTTCTAGATTTTCAAGATCAACTTCGACCAACCGCTTGGACCCCTGGTACACATTG 757
DB 662 GCCAGTTTCTGGAATTTCAAGACCAACTCGCTGCCCACTCGTGTGAGCTCTGTGTCCACATCC 721
QY 758 CGTTTAAACCGGAAGACGCTGCTTACAAAATATACCGGTATCTTTAAACATCGAAC 817
DB 722 AGTTGAAGCGGAAGAGCGGTCCAGTATAAACAAATCGCGGTGTCTTCAACATCGAGA 781
QY 818 GGGGGTTTTTCTTTGAACCCGATGAGAGCGCGGTGAGATTTAAATATATCGCATGAACACC 877
DB 782 AGGGTTCTTTTCGAGCCGATGAGAGGCTGTTGAATCAAGATCTGCGACGAACACC 841
QY 878 CGGCGTACAAATATGTTCCAGATGTCAGACGCGACGATGATGAGCATTCGTTTCGAAA 937
DB 842 CCGGGTACAGAAATATGATCCACCGGGGGCGACGCGCGCTGAGGAGCATTCCTTTCGAGA 901
QY 938 AAACCCAGATTTCCAAAGAGCCGGAACGCGGCTTCGGGCCCTGCTGAAAGAGACAATGC 997
DB 902 AGACGCAAGTTCTCTGAGAAAGCGGAGATCGCGGTCCGCAAGCTTCTGTCTGAAACGATGC 961
QY 998 CCCAGCTGGCAGACCGTCCATTTCAGCTTCGACGCAATTTGCTGGGTGTCGCGATACCGGA 1057
DB 962 CTACGTTGGGACCGGCGGTTTCAAGTTTTCGAAGGATCTGCTGGTGGGATACCCCA 1021
QY 1058 ATCGCGAATTCCTGATAGATTCGACATCCGAGTACCAAGTCTTGTGTTGGGCTGTGGT 1117


```
QY 408 CCGCCCGGAGCAATTTCTGTAACCTGGCCCCGG 439
Db 27185 GCTCACGGCGCCCTTCAACGACCTGGCGCGG 27216

RESULT 6
US-09-252-991A-7672
; Sequence 7672, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7672
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7672

Query Match 2.4%; Score 33.8; DB 4; Length 468;
Best Local Similarity 55.8%; Pred. No. 0.74;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 342 GGAGGCGCTGGATCGCTGGCGGTCCGGGTACGTCCGGCGGAGGATCCTTAATCTGGTGA 401
Db 251 GGATGGGCTCCAGCGGTTGGCGGCTTGCCTCTCTGGAGCAACCGATCAGCGCGTTGA 310

QY 402 ACTTACCCCGCCGAGCAATTTCTGTAACCTGGCCCCGGAAGGCGTGTGCAAGGTGA 458
Db 311 CCACCACCGCCAGGAGATTACCGCACTGTCCACCCAGAAAGCCGAGCGCGGTGA 367

RESULT 7
US-09-252-991A-7887/c
; Sequence 7887, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7887
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7887

Query Match 2.4%; Score 33.8; DB 4; Length 1032;
Best Local Similarity 55.6%; Pred. No. 1.3;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 342 GGAGGCGCTGGATCGCTGGCGGTCCGGGTACGTCCGGCGGAGGATCCTTAATCTGGTGA 401
Db 854 GGATGGGCTCCAGCGGTTGGCGGCTTGCCTCTCTGGAGCAACCGATCAGCGCGTTGA 795

QY 402 ACTTACCCCGCCGAGCAATTTCTGTAACCTGGCCCCGGAAGGCGTGTGCAAGGTGA 458
Db 311 CCACCACCGCCAGGAGATTACCGCACTGTCCACCCAGAAAGCCGAGCGCGGTGA 367

QY 408 CCGCCCGGAGCAATTTCTGTAACCTGGCCCCGG 439
Db 27185 GCTCACGGCGCCCTTCAACGACCTGGCGCGG 27216

RESULT 8
US-09-252-991A-7592
; Sequence 7592, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7592
; LENGTH: 2739
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7592

Query Match 2.4%; Score 33.8; DB 4; Length 2739;
Best Local Similarity 55.6%; Pred. No. 2.8;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 342 GGAGGCGCTGGATCGCTGGCGGTCCGGGTACGTCCGGCGGAGGATCCTTAATCTGGTGA 401
Db 2328 GGATGGGCTCCAGCGGTTGGCGGCTTGCCTCTCTGGAGCAACCGATCAGCGCGTTGA 2387

QY 402 ACTTACCCCGCCGAGCAATTTCTGTAACCTGGCCCCGGAAGGCGTGTGCAAGGTGA 458
Db 2388 CCACCACCGCCAGGAGATTACCGCACTGTCCACCCAGAAAGCCGAGCGCGGTGA 2444

RESULT 9
US-09-252-991A-7958/c
; Sequence 7958, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7958
; LENGTH: 2778
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7958

Query Match 2.4%; Score 33.8; DB 4; Length 2778;
Best Local Similarity 55.6%; Pred. No. 2.8;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 342 GGAGGCGCTGGATCGCTGGCGGTCCGGGTACGTCCGGCGGAGGATCCTTAATCTGGTGA 401
Db 418 GGATGGGCTCCAGCGGTTGGCGGCTTGCCTCTCTGGAGCAACCGATCAGCGCGTTGA 359

QY 402 ACTTACCCCGCCGAGCAATTTCTGTAACCTGGCCCCGGAAGGCGTGTGCAAGGTGA 458
Db 358 CCACCACCGCCAGGAGATTACCGCACTGTCCACCCAGAAAGCCGAGCGCGGTGA 302
```

```
RESULT 10
US-09-949-016-14624
; Sequence 14624, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 14624
; LENGTH: 147382
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(147382)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14624

Query Match      2.4%; Score 33.8; DB 4; Length 147382;
Best Local Similarity 52.5%; Pred. No. 53;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 406 ACCGCCGCGAGCAATTCGTAATCGCCCGGAGCGGTGTCGAAGTGATTTTCG 465
Db 47933 ACCTGGCAGATGAATTTGGTAGAGTGAATTTGGTAGGCTGTTGCAGTGAGCTAATGAA 47992

QY 466 GCTTGGAAAGGTAATTTGGCGCTTCGGCGCTGGGCGACATGCAAGATGCCCTTA 525
Db 47993 TGATGGAAGTGAGCAATGGGATCAGGTGTGGCGCATGACTTATCTGGAGATGTTG 48052

QY 526 GTGGCAGCAGCACCGGAAGCA 546
Db 48053 GCTGCAGAGAGGAGAGAAGGA 48073

RESULT 11
US-09-313-294A-4144/c
; Sequence 4144, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 4144
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700347784H1
US-09-313-294A-4144

Query Match      2.4%; Score 33.6; DB 4; Length 281;
Best Local Similarity 53.9%; Pred. No. 0.6;
Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
```

```
QY 14 GTGACGATGATGACCTGGCTCTCGCGTCACTAAGTCATCATCTCTCTGATCGTTG 73
Db 246 GTGAACAGCAGATGCTCTGCGGATGGCATAGGGTCATCACCTTCTCGCTGTCTGAG 187

QY 74 CCGGGACTTGGGGCACCTCAACGGCTCTGCACCTCGCGCGCGGATATACCAACGTTA 133
Db 186 CAGTGGCTTCAGCGTCGCCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 127

QY 134 CCGTGTCTG 141
Db 126 TCGTTCGTG 119

RESULT 12
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      2.4%; Score 33.6; DB 3; Length 4403765;
Best Local Similarity 59.4%; Pred. No. 2.9e+02;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 95 CGGCTCTGCACCTCGCGCGCGCGGATATACCAACGTTACCGTGTGCGAGCCCTATCTG 154
Db 835772 CCGAGCAGCAGCCCCCGCGCGCGGTACCCGACACCCCGCGGCGGCCCAAGCCG 835713

QY 155 TCCTAGCGCCATCTCCGCGGAAACGACGTGAACA 190
Db 835712 CCCTTACCGCGTTACCGCGCGGTTCGCGATCAGCCCA 835677

RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
```

US-09-103-840A-1

Query Match 2.4%; Score 33.6; DB 3; Length 4411529;
Best Local Similarity 59.4%; Pred. No. 2.9e+02;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 95 CGGCTCTGCACCTCGCGCGCGCGGATATACCAACGTTACCGTGTGGACCCCTATCCTG 154
Db 833594 CCGAGCAGACGCCCGCGCGCGCGGTACCAACGACACCCCGCGGACCCCAAGCCCG 833535
QY 155 TCCTAGCGCATCTCGCGCGGAAACGACGTGAACA 190
Db 833534 CCCTTACCGCGTTACCGCGGTTCGCGATCAGCCCA 833499

RESULT 14

US-09-252-991A-9143/c
; Sequence 9143, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9143
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9143

Query Match 2.4%; Score 33.4; DB 4; Length 801;
Best Local Similarity 48.2%; Pred. No. 1.5;
Matches 94; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
QY 62 TGATCGTTGGTGGCGGACCTTGGGGACCTCAACGGCTCTGCACCTCGCGCGCGCGGAT 121
Db 671 TGGTGATCGCGCGCGGCTTCCCGGGATCAACACCGCGCTTGAGCTGGCGGACGCGCA 612
QY 122 ATACCAACGTTACCGTCTGGACCCCTATCTCTCCCTAGCGCATCTCGCGCGGAAACG 181
Db 611 TCACGACATCGTGTCTGGAGCGCGCCACCTCGGCTTGGCGGTACCGCGCGCAACG 552
QY 182 ACGTGAACAAAGTCATTAGCAGTGGCCCAATATTTCGAATAACAAAGACGAAATCGAAGTGA 241
Db 551 GCGGCGAGATCATGGCGCGCATCGGCCACGACCTGGAGAGATCCGCAAGGACGTGGCG 492
QY 242 ATGAGATCTTGGCGG 256
Db 491 AGGACGCGCTGGCGG 477

RESULT 15

US-09-252-991A-8941
; Sequence 8941, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8941
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8941
Query Match 2.4%; Score 33.4; DB 4; Length 1818;
Best Local Similarity 48.2%; Pred. No. 2.8;
Matches 94; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
QY 62 TGATCGTTGGTGGCGGACCTTGGGGACCTCAACGGCTCTGCACCTCGCGCGCGCGGAT 121
Db 215 TGGTGATCGCGCGCGGCTTCTCCGGGATCAACACCGCGCTTGAGCTGGCGGACGCGCA 274
QY 122 ATACCAACGTTACCGTGTGGACCCCTATCTCTCTAGCGCATCTCGCGCGGAAACG 181
Db 275 TCACGACATCGTGTCTGGAGCGCGCCACCTCGGCTTGGCGGTACCGCGCGCAACG 334
QY 182 ACGTGAACAAAGTCATTAGCAGTGGCCCAATATTTCGAATAACAAAGACGAAATCGAAGTGA 241
Db 335 GCGGCGAGATCATGGCGCGCATCGGCCACGACCTGGAGAGATCCGCAAGGACGTGGCG 394
QY 242 ATGAGATCTTGGCGG 256
Db 395 AGGACGCGCTGGCGG 409
Search completed: May 29, 2005, 23:33:04
Job time : 277 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2005, 19:23:28 ; Search time 6433 Seconds
(without alignments)
10688.318 Million cell updates/sec

Title: US-10-622-893A-6
Perfect score: 1419
Sequence: 1 atggagggttcgggtgacga.....atccatcatcatcattaa 1419

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	846.4	59.6	1559	8 AFU82830	U82830 Aspergillus
2	685	48.3	1314	6 AR072866	AR072866 Sequence
3	685	48.3	1314	6 E14942	E14942 Aspergillus
4	685	48.3	1314	6 E14955	E14955 Aspergillus
5	685	48.3	1314	6 E27570	E27570 Thermostabl
6	685	48.3	1314	8 ATFAOA	Y09020 A. terreus m
7	657.6	46.3	1311	8 AB180733	AB180733 Aspergill
8	354.8	25.0	1526	8 AF035700	AF035700 Aspergill
9	350.4	24.7	1338	8 AB180732	AB180732 Aspergill
10	335.2	23.6	1320	6 BD173827	BD173827 Compositi
11	335.2	23.6	1320	6 E16562	E16562 Fusarium ox
12	333.6	23.5	1320	6 BD173829	BD173829 Compositi
13	332	23.4	1320	6 BD173828	BD173828 Compositi
14	332	23.4	1320	6 BD173830	BD173830 Compositi
15	127	8.9	110000	8 CR382133	Continuation (3 of
16	62.8	4.4	1314	6 AX712067	AX712067 Sequence
17	62.8	4.4	1314	8 AB116146	AB116146 Eupenicil
18	61.2	4.3	77986	8 EX897677	EX897677 Neurosp
19	57.6	4.1	1314	6 E22308	E22308 DNA encodin

20	57.6	4.1	1333	12	AF181866	AF181866 Synthetic
21	54.8	3.9	2780	8	AF416568	AF416568 Aspergill
22	53.6	3.8	1314	6	E22309	E22309 DNA encodin
23	53.6	3.8	1314	8	PJFAOP	Y09021 P.janthinel
24	50.4	3.6	1314	6	AX712065	AX712065 Sequence
25	50.4	3.6	1314	8	AB116147	AB116147 Coniocha
26	47.8	3.4	12770	8	SPAC139	AB132667 S.pombe c
27	44.8	3.2	2000	6	AX655393	AX655393 Sequence
28	44.2	3.1	125020	9	AF429315	AF429315 Homo sapi
29	43.2	3.0	43087	9	AP005155	AP005155 Homo sapi
30	43.2	3.0	110000	8	CR382138	CR382138 Debaryomy
31	40.6	2.9	479	14	AY180665	AY180665 Influenza
32	40.6	2.9	479	14	AY180666	AY180666 Influenza
33	40.4	2.8	125020	9	AF429315	AF429315 Homo sapi
34	40.2	2.8	190232	2	AC099532	AC099532 Bos tauru
35	38.8	2.7	3731	8	ENU28333	U28333 Emeritella
36	38.4	2.7	310174	1	AE016870	AE016870 Pseudomon
37	38.2	2.7	99427	2	AP003821	AP003821 Oryza sat
38	38.2	2.7	129320	2	AC151702	AC151702 Gallus ga
39	38.2	2.7	157797	8	AP005184	AP005184 Oryza sat
40	38.2	2.7	207960	10	AC124173	AC124173 Mus muscu
41	38	2.7	1246	6	AX164174	AX164174 Sequence
42	38	2.7	151986	8	AP005840	AP005840 Oryza sat
43	38	2.7	197358	8	AP004670	AP004670 Oryza sat
44	38	2.7	300684	1	AE017227	AE017227 Mycobacte
45	37.8	2.7	7218	6	166494	166494 Sequence 14

ALIGNMENTS

RESULT 1

AFU82830

LOCUS

DEFINITION

complete cds.

ACCESSION

U82830

VERSION

U82830.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Submitted (20-DEC-1996)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

Submitted (20-DEC-1996)

JOURNAL

FEATURES

source

CDS

AFU82830 1559 bp mRNA linear PLN 15-MAY-1997
Aspergillus fumigatus fructosyl amine: oxygen oxidoreductase mRNA,
complete cds.
U82830 GI:1769599
Aspergillus fumigatus
Aspergillus fumigatus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 1559)
Takahashi, M., Fischetrieder, M. and Monnier, V. M.
Molecular cloning and expression of amadoriase isoenzyme (fructosyl
amine:oxygen oxidoreductase, EC 1.5.3) from Aspergillus fumigatus
J. Biol. Chem. 272 (19), 12505-12507 (1997)

97284723
9139700
2 (bases 1 to 1559)
Takahashi, M., Fischetrieder, M. and Monnier, V. M.
Direct Submission
Submitted (20-DEC-1996) Pathology, Case Western Reserve University,
2085 Adelbert Rd., Cleveland, OH 44106, USA

Location/Qualifiers
1. .1559
/organism="Aspergillus fumigatus"
/mol_type="mRNA"
/db_xref="taxon:5085"
174..1490
/note="amadoriase II"
/codon_start=1
/product="fructosyl amine: oxygen oxidoreductase"
/protein_id="AAC49711.1"
/db_xref="GI:1769600"
/translation="MAVTKSSLLIVGACGTGCTSTALHLARQYNTVLDYPPVPSA
ISAGNDVKNVSISSQYSNNKDELEVNEILAEAFNGWKNDFPKFYVHDTGLMSACS
QEGDLRGVVRFGEDPNLVELTRPEQFKLAPEGVLQGFPGWKGKVFASGACMAH
RNALVAARQAQRMGVFTGTQGRVVTLLIFENNVDVKGATGDKGIWRABRTLCAG
ASAGQFLDKNQLRPTAHLVHALKEPRALYKNIPIVNIENIERGPFPEDEERGEIK
ICDEHGYTNVQSDGTWMSIPPEKTIQPKAEATRVALLKTMFOLADRPFSFARI
CWCAATANREFLDHRHPQYHSLVGGCAGSGRGFKILPSIGNLIVDAMEGKVPQKIHLE

ORIGIN		IKWNPDAIAARNRWRDLGRFGPNRVMDFDHVKEMTNVQYRDISKL"	
Query Match		59.6%;	Score 846.4; DB 8; Length 1559;
Best Local Similarity		76.8%;	Pred. No. 5.5e-240;
Matches 1036; Conservative		0;	Mismatches 316; Indels 0; Gaps 0;
QY	21	TGATGACCTGGCTCTCGCCGCTCACTAAGTCATCATCTCTCTGATCGTTGGTCCGGGAC	80
DB	161	TCAGGACACGTCATGGCGTAAACCAAGTCATCTTCCCTTTTGATCGTGGGGCAGGCAC	220
QY	81	TTGGGGCACCTCAACGGCTCTGCACTTCGGCCGCGGATATACCAACGTTTACCGTGT	140
DB	221	CTGGGGCACATCGACTGCTCTCCACCTTGGCAGCAAGAGGATACACAAATGTGACGGTCT	280
QY	141	GGACCCCTATCTGTCCTTAGCGCCATCTCCGCGGGAACGAGTCGACAAAGTCATTAG	200
DB	281	AGATCCCTACCCCGTTCCCTTCAGCCATCTCGGCTGGGAATGATGTGAACAGGTCACTC	340
QY	201	CAGTGGCCAAATATTCGAATAACAAAGACGAAATCGAAGTGAATGAGATCTTGGCGGAAGA	260
DB	341	CTCGGGCCATATAGCAACAAAGAGGAGGAAATGAGTCAACGAGATCTTGGCTGAGA	400
QY	261	GGCGTTTAAACGTTGGAAGAACCAACCGCTTTTCAAAACGTTATATCATGATACGGCCCT	320
DB	401	AGCGTTCAATGGCTGGAAGAAACGACCCCTTGTTCAAACCACTATATCAGGATACTGGATT	460
QY	321	GCTGATGTCGTGCTGTCGAGAGGGCCCTGGATCGGCTGGGGCTCGGGTAGCTCGGG	380
DB	461	GCTGATGTCGCGCTGCTCCAGGAAGGCTTGGACCGCTTGGAGTCCGTGTACAGGCCCGG	520
QY	381	CGAGGATCTTAATCTGTGTGAACCTTACCCGCGGAGCAATTTCTGTAACCTGGCCCCGGA	440
DB	521	TGAGGACCCCACTTGTGGAATCTGACACGCGCGGAGCAATTCGCAAAATAGCTCTGA	580
QY	441	AGCGGTGTTCAAGGTGAATTTCCGGGTGGAAGGTAATTTTCGCGGTTCCGCGCTGG	500
DB	581	GGGTGTTCTACAGGAGATTTCCCGGCTGGAAGGGCTACTTTGCGGTTTCAGGAGCTGG	640
QY	501	CTGGGCACATGCAAGGATGCTTAGTGAGCAGCAGCAGCGGAGCAGCAGCGCATGGTGT	560
DB	641	TTGGGGCCCATGCTCGCAATGCACTCGTGCTGTGCAAGGGAGGCTCAGAGAAATGGCGGT	700
QY	561	AAAAATTTGTTACTGGCACCCGCGAGGCTCGTGTAGTCACGTTAATCTTTGAAAAATAACGA	620
DB	701	GAAGTTCGTAATCTGGCACTCTCAGGSCAGAGTAGTCATCTATATTTGAGATTAACGA	760
QY	621	TGTAAAGGTGCCGTTACGGCGGATGGCAAAATTTGGAGAGCGGAACGTACATTCCTGTG	680
DB	761	TGTCAAGGTGCCGTTACCGGAGACGGCAAGATTTGGCGTGCAGAGCGCACATTCCTCTG	820
QY	681	TGCTGGGGCTAGCGCGGTCAGTTCCTAGATTTTCAGAAATCAACTTCGACCAACCGCTTG	740
DB	821	CGCGGGTGCAGGGCTGGTCACTTCCTGCACTTCAAGAAATCAGTTGGCTTCAACGGGATG	880
QY	741	GACCTGGTACACATTCGCTTAAAAACCGGAAGAACGTCGCTGTGTAACAAAATATACCGGT	800
DB	881	GAGCTGGTTCATATTGCTCTGAGGCTGAGGAGCGGGCTCTTTACAGAAATATCCAGT	940
QY	801	TATCTTTTAAACATCGAACCGGGGTTTTTTTGTGAACCGCATGAGAGCGCGTGCAGATTAA	860
DB	941	TATCTTCAACATTTGAGAGGGGTTCTTCTTTCGAACAGATGAGAGCGCGGTGAGATTAA	1000
QY	861	AATATGGCATGAACACCGGGCTACACAAATATGGTCCAGATGTCAGACGGCAGATGAT	920
DB	1001	GATCTGCGACGAAATCTCCGGGATATACCAATATGGTACAGTCTGCCGACGGCAGATGAT	1060
QY	921	GAGCATTCCTGCGAAAAACCCAGATTCCAAAAGAGCGAAACGCGCGTTCGGGCCCT	980
DB	1061	GAGCATTCCTTTTGAAGAACTCAGATTCTTAAAGAACCGGACGAGGGTTAGAGTCT	1120
QY	981	GCTGAAAGAGACAAATGCCAGCTGGCAGACCGGTCCATTGAGCTTCGCAACGCAATTTGCTG	1040

Db	1121	GCTTAAAGAGACGATGCCACAGCTTGCAGACCGTCCATTCAGTTTCGCCCAGGATTTGCTG	1180
QY	1041	GTGTGCCGATACCGGAATCGGGAATTCCTGTAGATGACATCCGACATCCGACGATACCACAGTCT	1100
Db	1181	GTGCGCGGACATGCCAACCGGAGTTCCTGATCGATCGCATCTCTAGTACCATTTGCT	1240
QY	1101	TGTGTTGGGCTGTGTGTGCGAGCGGAAGAGGGTTTAAATATCTGCTTCTATTTCGGGAATCT	1160
Db	1241	TGTGCTGGGCTGCGGCGCTTCGGCAGAGGATTCAAATATCTACCTTCAATTGGCAATCT	1300
QY	1161	CATTGTTAGCGGATGGAAGGTAAGTGGCGGCAAAAAATTCAGAAATTAATCAAGTGGAA	1220
Db	1301	CATCGTTGATGCTATGGAAGGCAAGGTCCCTCAAAAGATCCACGAATGATTTAAATGGAA	1360
QY	1221	CCCGGACATTTGGCGGAAACCGTAACCTGGCGTGATCTCTGGGCGGTTTTGGCGGTCCAA	1280
Db	1361	CCAGATATTGCTGCCAATCGCACTGGAGGATCTTTGGGAGATTCGGGGGTCCCA	1420
QY	1281	TGCTGTGATGGAATTTTCATGATGTGAAGGAATGGACCAATGTTTCAGTATCGTGATTTTC	1340
Db	1421	CAGAGTAATGGAATTCACGAGCTCAAGGAGTGGACAAATGTACAAATATAGATATTTTC	1480
QY	1341	CAAGCTGAAAGGAGAGTGTGAAGGTAAGCCAA	1372
Db	1481	CAAGTTATAAGTGATGTAGCAAAATCTGCCCAA	1512

RESULT 2

AR072866

LOCUS

Sequence 2 from patent US 5948659.

ACCESSION

AR072866

VERSION

AR072866.1 GI:9999629

KEYWORDS

Unknown.

SOURCE

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 1314)

AUTHORS

Kato, N., Sakai, Y., Tani, Y. and Fukuya, H.

TITLE

Recombinant fructosyl amino acid oxidase

JOURNAL

Patent: US 5948659-A 2 07-SEP-1999;

FEATURES

Location/Qualifiers

source

1..1314

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Query Match 48.3%; Score 685; DB 6; Length 1314;
Best Local Similarity 70.7%; Pred. No. 5.7e-192;
Matches 926; Conservative 0; Mismatches 380; Indels 3; Gaps 1;

QY	38	CGTCACTAAGTCATCATCTCTCTGATCGTTGTCGCGGAGCTTGGGGCACCTCAACGG	97
DB	5	CAGTCACCAAGTCTTCGTCGATATTGATCATCGGGCGGGCACCTGGGGTTGCTCAACTG	64
QY	98	CTCTGACCTTCGCGCGCGGATATACCAACGTTACCGTGTGAGCCCTATCTGTCC	157
DB	65	CCCTGATCTTGGCCGAGAGGATACACCAATGCTCACTGTCTTGCCTTACCCGCTACCGGTT	124
QY	158	CTAGCGCATCTCCGCGGAAACGAGTGAACAAAGTCATTAGAGTGGCCAAATATTCGA	217
DB	125	CATCAGCCATTTCCGCGCGGCAACGAGTCAACAAAGATCATCTGTCGCGGCAGTACAGCA	184
QY	218	ATAACAAAGCAAGTTCGAGTGAATGAGATCTTGGCGGAAGAGGGCTTTAAAGGTTGGA	277
DB	185	GCAAGAGGACGAGGTGGAAGTCAATGAGATTAATCGCCGAAACAGGCCCTTCAATGGCTGGA	244
QY	278	AGAACGACCGCTTTTCAAAACCGTATTATCATGATACGGGCCCTGCTGATGCTGCTTGCT	337
DB	245	AAATGACCCCATCTTCAGCCCGTACTACCAACGACCGGCGTGTGATGTCGCCACCA	304
QY	338	CGCAGAGGCGCTTGATTCGCTGGCGGTTCGGGTACGTCGCGGCGAGGATCTTAATCTGG	397

305 CACAGGAAGGATTGGAGCGTCTGGGGTCCGCGTGCACCTGAAGATGAACCCGATGTAG 364
Qy TGGAACTTACCGCCCGGAGCAATTCGTAACTGGCCCGGAGGGGTGTGCAAGGTG 457
Db CCGAAATTAAGTCTGGCCGCGAGCGATTCGCCAGCTGGCCCCC---GGCGTCTTGAAGGGTA 421
Qy ATTTTCGGGTGCGAAAGGTACTTTTCGGGTTCGCGGTTCGCGGTGGCGCATGCAAGGA 517
Db ACTTCCCGGTGGAGGGGTACACATTCCTCAACCGCGGTGGCGCATGCGCGCA 481
Qy ATGCTTTAGTGGCAGCAGCAGCGAAGCACAGCGCATGGGTGTAATAATTTTACTGGCA 577
Db ACGCCTTGGTCCGCGCGCGGAGGACACAGCGCTGGGTGCGTCTGTCGCGGAT 541
Qy CCGCGAGGTCGTGTAGTACAGTTATCTTTGAAATACAGATGTAAGAGTGGCGTTA 637
Db CGCGCAGGCGAGATCATCAGTTGATTTTGAAGCAACAGATGTGAAGGTGGCGTCA 601
Qy CGGCGATGCAAAATTTGAGAGCGCAAGCATATCTCTGTGTGGGTAGCGCG 697
Db CGGCGAGCGAAGATCTGGCGCGGAGAGATATCTCTGCGTGGTGGCGCGCG 661
Qy GTCAGTTCTTAGATTTCAAGATCAACTTCGACCAACCGCTTGGACCTCGGTACACATTG 757
Db GCCAGTTCTTGGATTTCAAGGACCAACTGCGTCCCACTGCGTGGACTCTGTCCACATCC 721
Qy CGTTAAACCGGAGAACGTCGTGTGTAACAAATATACCGGTATCTTTAAACATGCAAC 817
Db AGTTGAAGCGGAGAGCGTGCACAGTATTAACAAATCATCGCGTGTCTTCAACATCGA 781
Qy GGGGTTTCTTTTGAACCGATGAGGAGCGCGTGTGAGTTAAATATGCGATGACACC 877
Db AGGGTTCTTCTCGAGCGGATGAGGAGCGTGTGAAATCAAGATCTGCGACCAACC 841
Qy CGGCTACACAAATATGTCAGAGTGCAGACGCGCATGATGAGCATTCCTGTTCCGAA 937
Db CCGGTACAGATATGACCAACCGGGGCGGACGCGCGTGGAGGATTCCTTTCGAGA 901
Qy AATCCAGATTCGAAAGAGCGAAACGCGGTTCGGGCGCTGCTGAAAGAGACATGC 997
Db AGACGAGGTTCTCTCGAAGAGCGAGATGCGCGTCCGCAAGCTTCTGCTGAAACGATGC 961
Qy CCAGCTGGCAGACCGTCAATTCAGCTTCGACGATTCGCGTGTGGTGGCGGATACCGCA 1057
Db CTAGCTTGGGACCGCGCGTTCAGTTTCGAAAGGATCTGCTGTGTGGGATACCGCCA 1021
Qy ATCGCAATTCCTGATAGATCGACATCCGAGTACACAGTCTTGTGTGGGTGTGGTG 1117
Db ATCGAGTTTATCATTTGACCGTCAATCCGAAATACCGTGTGTTCTTGGTGTGGTG 1081
Qy CGAGCGAAGAGGTTTAAATATCTGCTTCTATTTGGGAATCTCATTTGACCGGATGG 1177
Db CTTTCAGGACGAGGTTCAAAATATCTTCCCTCGATCGGAAGCATCATCGCAGCGCCATGG 1141
Qy AAGTAAAGTCCGCAAAATTTACGAATTAATCAAGTGGAAACCGGACATTTGGCGGA 1237
Db AGGACAAACCCCGGCAAAATTCACAAAGTGTATCGCTGGAGCCCGGAAATCGCGATCA 1201
Qy ACCGTAACTGGCGTGTACTCTGGGGGTGTTTGGCGGTCCAAATCTGTGTGATGATTTTC 1297
Db ACCGTAACTGGGGGACAGATTAGTTCATTTGGAGGGCCCAACCGGTCATGATTTCA 1261
Qy ATGATGTGAAGGAATGGACCAATGTTTCAGTATCGTGTATTTTCCAAGCT 1346
Db ATGAAGTGAAGGAGTGGACTAATGTACCCCAAGGAGACATCTCGAAGTT 1310

RESULT 3

E14942
LOCUS Aspergillus terreus mRNA for fructosyl amino acid oxydase (FAOD-L).
DEFINITION E14942
ACCESSION E14942
VERSION E14942.1 GI:5709625

KEYWORDS JP 1998033177-A/1.
SOURCE Aspergillus terreus
ORGANISM Aspergillus terreus
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 1314)
AUTHORS Kato, N., Sakai, Y., Tani, Y. and Fukuya, H.
TITLE FRUCTOSYL AMINO ACID OXIDASE
JOURNAL Patent: JP 1998033177-A 1 10-FEB-1998;
KDK CORP
COMMENT OS Aspergillus terreus
PN JP 1998033177-A/1
PD 10-FEB-1998
PF 23-JUL-1996 JP 1996193344
PI KATO NOBUO, SAKAI YASUYOSHI, TANI YOSHIKI, FUKUYA HIROSHI PC
C12N15/09, C07H21/04, C12N1/21, C12N9/06, C12P21/02, C12Q1/26, PC
(C12N1/21,
PC C12R1:19), (C12N9/06, C12R1:19);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1. .1314
FT /strain= 'GPI' /organism= 'Aspergillus terreus' FT
FT CDS 1. .1314 /product= 'fructosyl amino acid oxydase'.
FEATURES
source 1. .1314
Location/Qualifiers
/organism= "Aspergillus terreus"
/mol_type= "genomic DNA"
/db_xref= "taxon:33178"

ORIGIN

Query Match 48.3%; Score 685; DB 6; Length 1314;
Best Local Similarity 70.7%; Pred. No. 5.7e-192;
Matches 926; Conservative 0; Mismatches 380; Indels 3; Gaps 1;
Qy 38 CCGTCACTAAGTCATCTCTCTGATCGTGGTGGGACCTTGGGGACCTCAACGG 97
Db 5 CAGTCACCAAGTCTCTGTCGATATTGATCATCGGGCGGCGACCTGGGGTGTCTCAATG 64
Qy 98 CTCTGACACTCGCGCGCGGATATACCAAGTTCCTGCTGGACCCCTATCTCTGCC 157
Db 65 CCCTGCTCTTGGCGCGAGGATACCAATGTCATCTGCTTACCCCGTACCGGTTC 124
Qy 158 CTAGCGCATCTCGCGGAAACGAGTGAACAAAGTCATTAGCAGTGGCCAAATTCGA 217
Db 125 CATCAGCATTTTCGCGGCAACGATCAACAGATCATCTGTCGCGCGCATCAGCA 184
Qy 218 ATAAACAAAGACGAAATCGAAGTGAATGAGATCTTGGCGGAGAGCGGCTTTAACGGTTGGA 277
Db 185 GCAAGAAGGACGAGTTCGAGTCAATGAGATTTATCGCCGAAACAGGCTTCAATGCGTGA 244
Qy 278 AGAAGACCGCGCTTTTCAACCGTATATCATGATACGGGCGCTGCTGATGCTGCTGCT 337
Db 245 AAAATGACCCCATCTTCAAGCGCTACTACCAAGACCGCGCGTGTGATGCTGCGCCACCA 304
Qy 338 CGCAGGAGGCGCTGATCGCTCGGGGTTCGGGTGCTGCGGCGGAGGATCTTAATCTGG 397
Db 305 CACAGGAAGGATTTGGAGCGTCTGGGGTTCGCGTTCGACCTGAAGATGAACCCGATGTAG 364
Qy 398 TGGAACTTACCCCGCGGAGCAATTTCTGTAATTCGGCCCGGAGGCGGTGTGCAAGGTG 457
Db 365 CCGAATTAAGTTCGCGCGGAGCATTCGCGCAGCTGGCCCCC---GGCGTCTTGAAGGGTA 421
Qy 458 ATTTTCGGGTGGAAGGGTACTTTGCGCGTTCGCGCGTCTGCTGGGCGCATGCAAGGA 517
Db 422 ACTTCCCGGTGGAGGGGTACCAATTCGCTCAAAACGCGGGCTGGCGCATGCGCGCA 481
Qy 518 ATGCTTAGTGGCAGCAGCAGCGGATGCGGATGCGGCTGTTAAATTTGTTACTGCGCA 577


```
Db 662 GCCAGTTTCTGGATTTCAGGACCAACCTGCTCCCACTGCTGGACTCTGTGTCACATCC 721
Qy 758 GGTAAACCCGAGAGAGCGTGTGTACAAAATATACCGGTATCTTTAACTGCAAC 817
Db 722 AGTTGAAGCCGGAAGAGCGTCCCACTATAAAACATGCCGTGTCTTCAACATCGAGA 781
Qy 818 GGGGGTTTTCTTTGAACCCGATGAGGAGCGGTGTAGATTAAATATGCGATGAACACC 877
Db 782 AGGGGTTCTTTCGAGCCGATGAGGAGCGTGTGTAAATCAAGATCTGCGACGAACACC 841
Qy 878 CGGGCTACACAAATATGTTCCAGATGCGAGACGCGCATGATGAGCATTCCTGTTGAAA 937
Db 842 CCGGGTACACGAATATGACACCGGGCGGACGCGCGTGAAGGACATTCCTTCGAGA 901
Qy 938 AAACCCAGATTCACAAAGAACCGAACCGCGGTTCCGGCCCTCTGCTGAAGAGACAATGC 997
Db 902 AGACGCGAGGTTCTTCGAGAGCGGAGATGCGCGTCCGCAAGCTTCTGTCTGAACGATGC 961
Qy 998 CCCAGCTGGCAGACCGTTCATTCAGCTTCGCAACGATTTGCTGTGTGCGCATACCGGA 1057
Db 962 CTGAGCTTGGGACCGGCGGTTTCAAGTTTCGCAAGATCTGCTGTGTGCGGATACCCCA 1021
Qy 1058 ATCCGGAATTCCTGATAGATCGACATCGCAGTACCAAGTCTTTGTGTGGGCTGTGGTG 1117
Db 1022 ATCCGAGTTTATCATTTGACCGTTCATCCGAATACCCGTCGCTTGTCTTGGGTGTGGTG 1081
Qy 1118 CGACCGAAGAGGGTTTAAATATCTGCTTCTTATTTGGGAATCTCATTTGTAACCGGATGG 1177
Db 1082 CTTGAGGACGAGGCTTCAAAATATCTTCCCTGATCGGAAGCATCATCGCAGACCCATGG 1141
Qy 1178 AAGGTAAAGTCCCGCAAAAATTCACGAATTAATCAAGTGAACCCGACATTCGGCGGA 1237
Db 1142 AGGACAAAACCCCGGCAAAAATTCACGAATTAATCAAGTGAACCCGACATTCGGCGGA 1201
Qy 1238 ACCGTAACTGCGGTGATCTCTGGGGGCTTTTGGCGGTCCAAATCGTGTGATGATTTTC 1297
Db 1202 ACCGTAACTGCGGGGACAGATTAGTGTGATTTGGAGGCGCCCAACCGGTCATGATTTCA 1261
Qy 1298 ATGATGTGAAGGAATGACCAATGTTGATGATGATGATGATGATGATGATGATGATGAT 1346
Db 1262 ATGAAGTGAAGGAGTGAGATTAATGTCACCCAAAGGAGACATCTCGAAGTT 1310

RESULT 5
E27570 LOCUS 1314 bp DNA linear PAT 18-JUN-2001
DEFINITION Thermotable fructosylamino-acid oxidase.
ACCESSION E27570
VERSION E27570.1 GI:13018233
KEYWORDS JP 1999221081-A/1.
SOURCE unidentified
ORGANISM Aspergillus terreus Gp1
REFERENCE 1 (bases 1 to 1314)
AUTHORS Yoshiki,T., Toru,K., Nobuyuki,Y., Akio,K. and Hiroshi,F.
TITLE Thermotable fructosylamino-acid oxidase
JOURNAL Patent: JP 1999221081-A 1 17-AUG-1999;
KYOTO DAIICHI KAGAKU CO LTD
COMMENT OS Aspergillus terreus Gp1
PN JP 1999221081-A/1
PD 17-AUG-1999
PF 10-FEB-1998 JP 1998028169
PR
PI YOSHIKI TANI,TORU KATSURAGI,NOBUYUKI YOSHIDA,AKIO KATO, PI
HIROSHI FUKUIE
PC C12N15/09,C07K19/00,C12N1/21,C12N9/06/(C12N15/09,C12R1:66),
PC (C12N1/21,C12R1:19),(C12N9/06,C12R1:19),C12N15/00,(C12N15/00,
PC C12R1:66)
CC Strandedness: Double;
CC Topology: Linear;
PH Key
FT Location/Qualifiers
FT source 1..1314
/organism='Aspergillus terreus Gp1'.
```

```
FEATURES
source
Location/Qualifiers
1..1314
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 48.3%; Score 685; DB 6; Length 1314;
Best Local Similarity 70.7%; Pred. No. 5.7e-192;
Matches 926; Conservative 0; Mismatches 380; Indels 3; Gaps 1;
Qy 38 CCGTCACTAAGTCAATCATCTCTGATCGTGTGGTGGCGGACCTTGGGGCACCTCAACGG 97
Db 5 CAGTCACCAAGTCTTCTGATATTGATCATCGGGCGGGCCACTTGGGGTGTCTCACTG 64
Qy 98 CTCTGCACCTCGCGCGCGCGGATATACCAAGTTTACCGTCTGGACCCCTATCTCTGTC 157
Db 65 CCCTGCATCTTGGCCGCGAGAGGATACACCAATGTCACTGTCTTGGACCGTACCGGTT 124
Qy 158 CTAGCGCATCTCGCGCGGAAACGACGTGAACAAAGTCAATTAGCAGTGGCCAAATTCGA 217
Db 125 CATCAGCCATTTCCGCGCGCAACGACGTCAACAAAGATCATCTCGTCCGGCCAGTACA 184
Qy 218 ATAAACAAGACGAATCGAAGTGAATGAGATCTTTGGCGGAAGAGCGGTTTAAACGTT 277
Db 185 GCAGAAGAGAGGTGAAAGTCAATGAGATTATCGCCGAACAGGCTTCAATGGCTGGA 244
Qy 278 AGAACGACCCGCTTTTCAAAACCGTATTATCATGATACGGGCTCTGATGTCTGTGCT 337
Db 245 AAATGACCCCATCTTCAAGCGGTACTACCAACGACACCGGCGTGTGATGTCCGCCA 304
Qy 338 CGCAGGAGGCTGTGATCGGCTGCGGTGACGTCGCGGCGAGAGATCCTTAATCTGG 397
Db 305 CACAGGAAGATTGGAGCGTCTGGGGTCCGCGTGCACCTGGAAGATGAACCCGATGTAG 364
Qy 398 TGAACCTTACCGCGCGGAGCAATTTCTGTAACCTGCGCCCGGAAGGCGTGTGCAAGTG 457
Db 365 CCGAATTGATCGGCGCGGAGCAGTTCCGCCAGCTTGGGCCCCC---GGCGTCTTGAAGG 421
Qy 458 ATTTTCCGGTTGGAAAGGGTACTTTTTCGGGCTTCCGGCGTCTGGGCGACATGCAAGGA 517
Db 422 ACTTCCCGGTTGGAGGGGTACCACATTCGCTCAACGCGGGCTGGGCGCATGCGCGCA 481
Qy 518 ATGCTTAGTGGCAGCAGCAGCGAGCAGCAGCGCATGGGTGTAAATTTGTTACTGGCA 577
Db 482 ACSCCTGTGTGCGCGCGCGCGGAGGACAGCGCTGGGTGTGCGCTTCGTCGCGGAT 541
Qy 578 CCGCGCAGGGTCTGTGTAGTCACTGTTAATCTTTGAATAACGATGTAAGGTCGCGTTA 637
Db 542 CGCGCAGGGCAGAGTCATCACGTTGATTTTGAAGAACACGATGTGAAGGGTCCCGTCA 601
Qy 638 CGGCGCATGGCAAAATTTGGAGAGCGGAACGTACATTTCTGTGTCTGGGGCTAGCGCG 697
Db 602 CGCGGACGCAAGATCTGCGGGCGGAGCAGACTATCTCTGCGCTGTGGCGCCCGCG 661
Qy 698 GTCACTTCTAGATTTCAGAAATCAACTTCGACCAACCGTGTGGACCTGTGACACATTTG 757
Db 662 GCCAGTTTCTGGATTTCAGGACCACTGCGTCCCACTGGGTGAGCTCTGGTCCACATCC 721
Qy 758 CGTTAAACCGGAAGACGTCGTTGTACAAAATATACCGGTTATCTTTAAACATCGAAC 817
Db 722 AGTTGAAGCCGGAAGAGCGTGCACGATATAAAACATGCGCGGTGGTCTTCAACATCGA 781
Qy 818 GGGGGTTTTTCTTTTGAACCCGATGAGGAGCGGTGAGATTAAATATGCGATGAACACC 877
Db 782 AGGGGTTCTTCTCGAGCCGATGAGGAGCGTGGTGAATCAAGATCTGCGACGAACACC 841
Qy 878 CGGGCTACACAAATATGTTCCAGATGCGAGTGAACGCGCAGATGATGAGCATTCCTGTT 937
Db 842 CCGGGTACACGAATATGACCAACCGGGCGGACGCGCGGTGAGGAGCATTCCTCTTCGA 901
Qy 938 AAACCCAGATTCCAAAGAGCCGAACGCGGCTTGGGGCCCTCTGCTGAAGAGACAATGC 997
```

Db 902 AGACGAGGTTCTCTCGAGAAGCGAGATCGCGTCCGCAAGCTTCTGTCTGTAACGATCG 961
QY 998 CCAGAGTGGCAGACCGTCCATTCCAGCTTCGCGACGATTTGCTGTGCGGATACCGCGCA 1057
Db 962 CTGAGCTTGGGACCGCGCTTCAGTTTTCGAAAGATCTGCTGTGCGGATACCCCCA 1021
QY 1058 ATCGCGAATTTCTGTAGATCGACATCCGAGTACACACAGTCTTGTGTTGGGCTGTGGTG 1117
Db 1022 ATCGCGAGTTTATCATGACCGTTCATCCGATACCCGTCGCTTGTCTTGGTGTGGTG 1081
QY 1118 CGAGCGAGAGAGGTTTAAATATCTCGCTTCATTCGGAATCTCATTTGTTGAGCGCATGG 1177
Db 1082 CTTGAGGAGGAGGTTTCAAATATCTTCCCTCGATCGAAGGATCATCGCAGACGCCATCG 1141
QY 1178 AAGTAAAGTGGCGCAAAAATTCAGAAATTAATCAAGTGAAGTGAACCGGACATTCGCGCA 1237
Db 1142 AGGACAAAACCCCGGCAAAAATTCAGAGTGAATCGCTGGAGCCCGGAAATCGCGATCA 1201
QY 1238 ACCGTAACCTGGCGTGATCTCTCGGCGCTTTTGGCGGTCCAAATCGTGTGATGATTTTC 1297
Db 1202 ACCGTAACCTGGGCGGACATTTAGTTCGATTTGGAGGCCCAACCGGTCATGATTTCA 1261
QY 1298 ATGATGTGAAGAAATGGACAAATGTTTCAGTATCGTGATATTTTCAAGCT 1346
Db 1262 ATGAAGTGAAGAGTGGACTTAATGTCAACCAAGGAGACATCTCGAAGTT 1310

RESULT 6

ATFAOA 1314 bp mRNA linear PLN 16-APR-1997
LOCUS A.terreus mRNA for fructosyl amino acid oxidase.
DEFINITION Y09020
ACCESSION Y09020.1 GI:1653969
VERSION faoA gene; flavoprotein; fructosyl amino acid oxidase.
KEYWORDS Aspergillus terreus
SOURCE Aspergillus terreus
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 Yoshida,N., Sakai,Y., Isogai,A., Fukuya,H., Yagi,M., Tani,Y. and
AUTHORS Kato,N.
TITLE Primary structures of fungal fructosyl amino acid oxidases and
their application to the measurement of glycosylated proteins
JOURNAL Eur. J. Biochem. 242 (3), 499-505 (1996)
MEDLINE 97175014
PUBMED 9022674
REFERENCE 2 (bases 1 to 1314)
AUTHORS Kato,N.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1996) N. Kato, Fac. of Agriculture, Kyoto Univ.,
Kitashirakawa-oiwake-cho, Sakyo-ku, Kyoto 606-01, JAPAN

FEATURES

source
1. .1314
/organism="Aspergillus terreus"
/mol_type="mRNA"
/isolate="GPI"
/db_xref="taxon:33178"
/cell_line="E.coli SOLR"
1. .1314
/gene="faoA"
1. .1314
/gene="faoA"
/codon_start=1
/product="fructosyl amino acid oxidase"
/protein_id="CAA70218.1"
/db_xref="GI:1653970"
/db_xref="GOA:Q92204"
/db_xref="UniProt/Ts:EMBL:Q92204"
/translation="MPVTKSSSILLIGAGTWGCGSTALHLARRGYTNVTLDPVPSA
ISAGDNVKLIISGQSSKXKDEVEINEITAEAFNWKNDPIFKPYHDTGVMSAITT
QEGRLGRVREDFPDLAPFLQAPLKGPNFPGWRGYHRSNAGWAHAR
NALVAAREAQRLGVRFVAGSPQGRVITLIFENNVDKAVTADGKIWRABQTLICAGA
AAGQFLDFKDLRPTAWTLVHILQKPEERAQYKKNMPVVFVIEKGFPEPDERGEIKI

ORIGIN

Query Match 48.3%; Score 685; DB 8; Length 1314;
Best Local Similarity 70.7%; Pred. No. 5.7e-192;
Matches 926; Conservative 0; Mismatches 380; Indels 3; Gaps 1;
-
QY 38 CGTCACTAAGTCATCATCTCTCTGATCGTTGGTCCGCGGACTTGGGCGACCTCAACGG 97
Db 5 CAGTCAACCAAGTCTTCTGTCGATATTGATCATCGGGCGGGCACCTGGGGTGTCTCAACTG 64
QY 98 CTCTGCACTCCGCGCGCGCGATATACCAACGTTTACCGTGTGACACCCCTATCTCTGTCC 157
Db 65 CCTGTGATCTTCCCGCAGAGGATACACCAATGTCACTGTCTTGAACCGTACCCGGTTC 124
QY 158 CTAGCGCCATCTCCGCGGAAACGAGTGAACAAAGTCAATTAGACGTGCGCCAAATATCGA 217
Db 125 CATCAGCCATTTCCGCGCGCAACGACGTCACAAAGATCATCTCTGTCGCGCCAGTACAGCA 184
QY 218 ATAAACAAGACAAATCGAAGTGAATGAGATCTTGGCGGAAGAGCGGTTTAAAGGTTGA 277
Db 185 GCAAGAAGACGAGGTGCAAGTCAATGAGATTATCGCCGAACAGGCTTCAATGGCTGA 244
QY 278 AGAACGACCCGCTTTTCAAACCGTATTATCATGATACGCGCCCTGCTGATGTCTGTGTGCT 337
Db 245 AAATGACCCCATCTTCAAGCCGTACTACACGACACCGCGCTCGTGATGTCCGCCACCA 304
QY 338 CGCAGAGGCGCTGATTCGCTGGGCTCCGGGTAGCTCCGGGCGAGGATCTTAATCTGG 397
Db 305 CACAGGAAGGATTGGAGCGTCTGGGGTCCGGTGCAGCTGAAAGATGAACCCGATGTAG 364
QY 398 TGAATCTACCGCGCGGAGCAATTCGTAACCTGCCCCCGGAGCGGTGTGCAAGGTG 457
Db 365 CCGAATGATCTCGCGCGGAGCAGTTCGCCAGCTGGCCCCC---GGCGTCTTGAAGGGTA 421
QY 458 ATTTTCCGGGTTGGAAGGGTACTTTTGGCGCTTTCGGCGCTGGCTGGGCACATGCAAGGA 517
Db 422 ACTTCCCGGTTGGAGGGGTACCACTTCGCTCAACCGGGGCTGGGCGCATGCCGCA 481
QY 518 ATGCTTTAGTGGCAGCAGCACCGAAGCAGCGCATGGGTGTAAATTTGTTCTGGCA 577
Db 482 AGCCCTGTGTCGCGCGCGCGGAGGCAAGCGCTGGGTGTGCGCTTCGTCGCGGAT 541
QY 578 CCGCGCAGGCTGTGTAGTACAGTTAATCTTTGAAATAACGATGTAAAGGTGCGCTTA 637
Db 542 CGCGCGAGGCGAGGTCATCAGCTTGATTTTGAAGAACCAACGATGTGAAGGGTGCGCTCA 601
QY 638 CGGGCGATGGCAAAATTTGGAGAGCGGAACGTACATTCCTGTGTGGGGCTAGCGCGG 697
Db 602 CGGCGACGCGAAGATCTGGCGGCGGAGAGGATATCTCTCGCTGGTGGCGCGCGCG 661
QY 698 GTCAGTTCCTAGATTTCAAGAATCAAATTCGACCAACCGCTTGGACCCCTGGGTACATTTG 757
Db 662 GCCAGTTTCTGGATTTCAAGGACCAACTGCTGCCACCTGCTGGTGGTCTGGTCCACATCC 721
QY 758 CGTTAAACCGGAGAAACGTGCTGTACAAATAATACGGTTATCTTTTAACATCGAAC 817
Db 722 AGTTGAAGCGGAAGAGCGTCCAGTATAAAACATGCGCGGTGCTTCAACATCGAGA 781
QY 818 GGGGGTTTTTCTTTGAACCCGATGAGGAGCGCGGTGAGATTAAATATGCGATGAACACC 877
Db 782 AGGGGTCTTCTTCGAGCGCGGATGAGAGGGTGTGAAATCAAGATCTGCGACGAACACC 841
QY 878 CGGGCTACACAAATATGGTCCAGAGTGCAGACGCGCAGATGATGAGCATTTCCGTTGCAAA 937
Db 842 CCGGGTACAGAAATATGACCAACCGGGGCGGACGCGCGCGTGAAGAGCATTTCCCTTCGAGA 901
QY 938 AAACCCAGATTCCAAAGAGACGGAACCGCGCTTGGGCGCTTGGGCGCTGTAAGAGACATGC 997
Db 902 AGACGAGGTTCTCTCGAGAAGCGGAGATGCGCGCTCCGCAAGCTTCTGTCTGAAACGATGC 961

```

Qy 998 CCACGCTGGCAGACCGTCCATTGAGCTTCGACGCAATTTGCTGGTGGCCGATACCGGA 1057
Db 962 CTCAGCTTGGCGACCGCGCGTTCAGTTTCGCAAGGATCTGCTGGTGGCGGATACCGCCA 1021
Qy 1058 ATCCGGAATTCCTGATAGATCGACATCCGACGATACCAAGTCTTGTGTGGGCTGTGGTG 1117
Db 1022 ATCCGGAATTCATGACGCTCATCCGGAATACCGTCCGCTGCTTGTGTGGTGTGGTG 1081
Qy 1118 CGAGCGAAGAGGTTTAAATATCTGCTTCTATTTGGGAATCTCATTTGTTGACGCGATGG 1177
Db 1082 CTTGAGCAGCAGGCTTCAAAATATCTTCCCTGATCGGAAGCATCATCGCAGACCCATGG 1141
Qy 1178 AAGTAAAGTCCCGCAAAAATTCACGAATTAATCAAGTGAACCCCGACATTTGCGCGGA 1237
Db 1142 AGGACAAAACCCCGCAAAAATTCACGAATTAATCAAGTGAACCCCGACATTTGCGCGGA 1201
Qy 1238 ACCGTAAGTCCCGCAAAAATTCACGAATTAATCAAGTGAACCCCGACATTTGCGCGGA 1297
Db 1202 ACCGTAAGTCCCGCAAAAATTCACGAATTAATCAAGTGAACCCCGACATTTGCGCGGA 1261
Qy 1298 ATGATGTAAGGAATGACCAATGTTCAAGTATCTGATATCTGATATTTTCAAGCT 1346
Db 1262 ATGAAGTGAAGGAGTGAATTAATCAAGTATCTGATATCTGATATTTTCAAGCT 1310

RESULT 7
AB180733
LOCUS
DEFINITION Aspergillus oryzae FAOAO2 mRNA for fructosyl-amino acid oxidase,
complete cds.
ACCESSION AB180733
VERSION AB180733.1 GI:53850451
KEYWORDS
SOURCE Aspergillus oryzae
ORGANISM Aspergillus oryzae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1
AUTHORS Akazawa,S., Karino,T., Yoshida,N., Katsuragi,T. and Tani,Y.
TITLE Functional Analysis of Fructosyl-Amino Acid Oxidases of Aspergillus
oryzae
JOURNAL Appl. Environ. Microbiol. 70 (10), 5882-5890 (2004)
PUBMED 15466528
REFERENCE 2 (bases 1 to 1311)
AUTHORS Akazawa,S., Karino,T., Yoshida,N., Katsuragi,T. and Tani,Y.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2004) Shin-ichi Akazawa, Graduate School of
Biological Sciences, Nara Institute of Science and Technology,
8916-5 Takayama, Ikoma, Nara 630-0192, Japan
(E-mail: s-akazawa@bs.ais-t-nara.ac.jp, Tel:81-743-72-5423,
Fax:81-743-72-5429)

FEATURES
source
    1..1311
        /organism="Aspergillus oryzae"
        /mol_type="mRNA"
        /db_xref="taxon:5062"
    1..1311
        /gene="FAOAO2"
    1..1311
        /codon_start=1
        /gene="FAOAO2"
    /product="fructosyl-amino acid oxidase"
    /protein_id="BAD54825.1"
    /db_xref="GI:53850452"
    /translations="MTVAKSSSILIGAGTWAGTALHLGRGVNTVLDPVTPSA
    ISAGDVNKLIISSQVSNKDEIEVNEILAEAFKWTDPDLPKPYHDGVMSACS
    SAGDRUGIRVEEEDPVSEVTKPHFROLAPVAVLKNFPNGMRGHIRSNAGWAHAR
    NALVAIAREAEKLGKVFVTQGRVITLI FENNDRKGVAVTADGKI WRAEQTIVLCAGN
    AAQFLDFKQRLTAMTFLAHLRKPEERALKYKLPVFNTEKGFPEPDERGEIKIC
    DEHPGTNMVMSADGHLTSLPFEKTOIPKSEARVRLALSETWPLQADRPSPFARVCW
    CADTANREFIIDRHPHPSLVLCGASGRGFKYLPISIGNLIIDVAIEDKVEKVKLTR
    WSPDIAVDRKWRDTLGRFGPNRUMDFHVKWNTVQNKDTAKL"

```

ORIGIN

```

Query Match 46.3%; Score 657.6; DB 8; Length 1311;
Best Local Similarity 69.9%; Pred. No. 8.1e-184;
Matches 917; Conservative 0; Mismatches 389; Indels 6; Gaps 2;

Qy 38 CCGTCACTAAGTCATCATCTCTCTGATCGTTGGTCCGGGACTTTGGGGACCTCAACGG 97
Db 5 CTGTGCGCAAAATCTTCTCAATCTGATCATTCGCGCAGGACATTTGGGGCGCTTCAATG 64
Qy 98 CTCTGCACCTCGCGCGCGCGGATATACCAACGTTACCGTGTGGACCCCTATCTGTGCC 157
Db 65 CCCTTCACTTGTGTCGACAGGATACACCAATGTCAACGCTTAGACCTTTACACAGTGC 124
Qy 158 CTAGCGCATCTCGCGCGGAAACGAGTGAACAAAGTCAATAGCAGTGGCCAAATATTCGA 217
Db 125 CCTCAGCAATTTTCAGCTGGAAATGACGTGAACAAGATCATCTCTCGGGGCAATACAGA 184
Qy 218 ATAAACAAAGACGAAATCGAAGTGAATGAGATCTTTGGCGGAAGAGCGGTTTTAACGGTTGA 277
Db 185 ACAAAGAGATGAGATGAAAGTTAAAGAAATTCGCGCAGGAGGCAATTCAGAGCTGGA 244
Qy 278 AGAACGACCCGCTTTTCAAAACCGTATTATCATGATACGGGCTGCTGATGTCTGCTTGC 337
Db 245 CAACCGACCTTTGTTCAAGCCATCTACTACACGACACTGGGGTGTGTAATGTCTGCTTGA 304
Qy 338 CGCAGGAGGCTGTGATCTGCTGGCGTCCGGGTACGTCGCGGCGAGGATCTTAATCTGG 397
Db 305 GCAGCGCGGCTGTGGATCTCGAATTCGAGTAAAGCGCGGAAGAGGAACTCTGATGTTT 364
Qy 398 TGGAACTTACCCCGCGGAGCAATTTGTAACCTGGCCCGGAGGGGTGTGCAAGGTG 457
Db 365 CCGAAGTCAGAAAGCGGAGCACTTCGCCCACTGGCCCCCGCT--GTGCTGAAGAGAA 421
Qy 458 ATTTTCGGGTTGGAAGGGTACTTTTCGCGCTTCGCGCTGGCTGGGCGACATCAAGGA 517
Db 422 ACTTCGCGGGTGGAGAGGCTACCACTTCGTTGCAACGCTGGCTGGCGCACGCCGAA 481
Qy 518 ATGCTTTAGTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 577
Db 482 ATGCTCTCTGGCGGCTTATACGCGAAGCAGAGAACTTGGTGTAAATTCGTAAACGCA 541
Qy 578 CCGCGAGGCTGTGTAGTCACTGTTTCTTGAATAACGATGTAAGAGTGGCGCTTA 637
Db 542 CC---CAAGGAAGAGTCACTCACCTTATCTTCGAGAACACAGCTCAAGGGCGCAGTCA 598
Qy 638 CGGCGCATGGCAAAATTTTGAGAGCGGAACTGATCATCTCTGTGTGGTGGGCTAGCGCG 697
Db 599 CCGCGCAGGAAGATCTGGCGCGGAGCAACAGTCTCTGCGCTGGCGCAATGCTG 658
Qy 698 GTGATTTCTAGATTTCAAGAAATCAACTTCGACCAACCGCTTGGACCCCTGTTACACAT 757
Db 659 CGCAGTTCTTGGATTTTAAAGGACAGCTCGCGCCCAACGCGATGGACACTCGCCCATAT 718
Qy 758 CGTTAAACCGGAAGAACGCTGCTGTGTACAAAATATACCGGTTATCTTTAAACATCGAAC 817
Db 719 GGCTCAAACTGAGGAACCGCGCTCTACAAAATCTTGCCTGGTGAATTTCAACATAGGA 778
Qy 818 GGGGGTTTTCTTTTGAACCGATGAGAGCGCGGTGAGATTAATAATATGCGATGAACACC 877
Db 779 AAGGATTTTCTTCGAGCCTGATGAGAGCGCGGGAGATCAAGATCTGCGACGAACATC 838
Qy 878 CGGGCTACACAAATATGGTCCAGATGTCAGACGCGCATGATGAGCATTCCTGTTCAAAA 937
Db 839 CGGGATACACTAACATGTTTAAATCTCGCGATGGCCACTTGAACGAGTTTGGCCCTTTGAGA 898
Qy 938 AAACCCAGATTTCCAAAAGAGCGGAACGCGCTTCGGGCCCTCTGTAAGAGCAATATGC 997
Db 899 AGACCCAGATTTCCCAAAGAGTCTGAAAGTGAAGTCTAGAGCTTACTATCGGAGACCAT 958
Qy 998 CCGAGCTGGCAGACCGTCCATTTCAGCTTTCGACGCAATTTCTGTGTGGCGCATACCGCA 1057
Db 959 CTCAATTAGCCGATCGCCCATTTAGCTTTCGCCCGGCTTCTGCTGGTGTGGGACACCGCA 1018

```



```
Db 1213 GAGTAACACAGAGGAGTGAAGACATCGTTTCGATGGAGCCACAGACGGCTGTGCA 1272
Qy 1239 CCGTAACCTGGCGTCACTCTGGGGCGTTCGGGGTCCAAATCGTGTGATGATTTTCA 1298
Db 1273 TCGAGATTGAGACGCACTCAGATCGCTTTGGCGGCGCTGACAGGATCATGGATTTTCA 1332
Qy 1299 TGATGT 1304
Db 1333 GCAGGT 1338

RESULT 9
AB180732 1338 bp mRNA linear PLN 07-OCT-2004
LOCUS Aspergillus oryzae FAOa01 mRNA for fructosyl-amino acid oxidase,
DEFINITION complete cds.
ACCESSION AB180732
VERSION AB180732.1 GI:53850449
KEYWORDS
SOURCE Aspergillus oryzae
ORGANISM Aspergillus oryzae
Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1
AUTHORS Akazawa,S., Karino,T., Yoshida,N., Katsuragi,T. and Tani,Y.
TITLE Functional Analysis of Fructosyl-Amino Acid Oxidases of Aspergillus
oryzae
JOURNAL Appl. Environ. Microbiol. 70 (10), 5882-5890 (2004)
PUBMED 15466528
REFERENCE 2 (bases 1 to 1338)
AUTHORS Akazawa,S., Karino,T., Yoshida,N., Katsuragi,T. and Tani,Y.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2004) Shin-ichi Akazawa, Graduate School of
Biological Sciences, Nara Institute of Science and Technology,
8916-5 Takayama, Ikoma, Nara 630-0192, Japan
(E-mail:s-akazawa@bs.ais.t-nara.ac.jp, Tel:81-743-72-5423,
Fax:81-743-72-5429)

FEATURES
source
1. .1338
/organism="Aspergillus oryzae"
/mol_type="mRNA"
/db_xref="taxon:5062"
1. .1338
/genes="FAOa01"
/codon_start=1
/product="fructosyl-amino acid oxidase"
/protein_id="BAD54824.1"
/db_xref="GI:53850450"
/translations="MTSSKLTPTSSILIVGATGWCSTALHLARRGKYNVLDPHPV
PSPIAAGNDINKMEHREVKASPTCTRAALKGWNKDPVFPYFHTGAI
VSGHTASLIKHIOHEIDSDAGFIKLNTAEFRKTPPGILTGNFPKWLKNGA
GWIHAKMDSAYTEAKRLGVFTITGSPGDVSVLIVGVGAVGATGDTVHRADHT
ILSAGMSDRLLPFKQLRPTANTILCHIRMTPEAKYKRLNPLVFNFAVGFPMEPDE
NHEKICDEHPKFNVPDPKHGEVRSIPFAKHQIPLEABARDFLRDTPHPLADR
PLSPFARICWADVTDFRALFDHREYRSLSLVAGSGNGAMQPTTIGGTFADALEGL
QKELKHALRWRPETAQRDKWDTQNRFGGPNKVMDFQKUGENETWIKIGKSRL"

ORIGIN
Query Match 24.78; Score 350.4; DB 8; Length 1338;
Best Local Similarity 55.94; Pred. No. 2.1e-92;
Matches 713; Conservative 0; Mismatches 551; Indels 12; Gaps 2;

Qy 41 TCACTAAGTCATCATCTCTCTGATCGTTGTCGCGGACCTTGGGGCACCTCAACGGCTC 100
Db 17 TGACTCCACATCATCTATCTTAATGTTCGTGCGAGGACCTGGGGTGTCTACTGCTT 76
Qy 101 TGCACCTCGCGCGCGGATATACCAACGTTACCGTGTGGACCCCTATCTGTCCCTA 160
Db 77 TACATCTTGGCCGTCGAGGATACAAAATGTACGGTCTTAGATCCGCACCCGGTCCCTT 136
Qy 161 GCGCCATCTCGCGGGGAAACGAGTGAACAAAGTCAT-----TAGCAGTGGCCATATT 214
```

```
Db 137 CTCCCATTCAGCTGGCAATGACATTAACAAGATTATGGAGCACAGGAGTAAGAAGCTT 196
Qy 215 CGAATAACAAGACGAATCGAAGTGAATGAGATCTTTGGCGGAAGAGCGGTTTAAACGGTT 274
Db 197 CTGAAACCGATCTCTTGAGTATGCGCTTCTCAAATGACACACAGAGCTGCACGAAAGGTT 256
Qy 275 GGAAGAACGACCCGCTTTTCAAACCGTATTATCATGATACGGGCTGCTGATGCTGCTT 334
Db 257 GGAATAACGACCCAGTATTCAGCCATATCTCCATGAACCGGGGCAATAGTTTCTGGCC 316
Qy 335 GCTCGCAGGAGGGCTCGATCGCTGGGCTCGGGTACGTCGCGGAGGATCCTAATC 394
Db 317 ACACCGCTCTTTGATTAAACATATACAGAACAGAAATCGACTGTCAGACCGCGAGT 376
Qy 395 TGGTGGAACTTACCGCCCGGAGCAATTTTCGTTAACTTGGCCCGGAAAGCGTGTGTCAAG 454
Db 377 TCATAAAATTTGAACACCGCAGAGGATTTCCGCAAAACTATGCCCCCGGAATCCTCACTG 436
Qy 455 GTGATTTTCCGGGTTGAAAAGGTAATTTCGTTAACTTGGCCCGGAAAGCGTGTGTCAAG 514
Db 437 GCAACTTTCCCGGCTGGAAGGCTGGCTGAACAAGACCGGCGCGGATGGATCCACGCCA 496
Qy 515 GGAATGCTTAGTGGCAGCAGCAGCGAAGCAGCGCATGCGGTGTAATAATTTGTTACTG 574
Db 497 AGAAGGCCATGTTCTCCGATACACCGAAGAAAGCCCTAGGAGTCACTTTCAATCACG 556
Qy 575 GCACCCCGCAGGCTGCTGTAGTCACTTAATCTTTGAAAATAACGATGTAAGAGGTGCG 634
Db 557 GCTCCCTCGAAGGAGAGCTTGTATCTCTAATTTACGAGATGGAGACGTAGTCGAGCCA 616
Qy 635 TTACGGCGGATGGCAAAATTTGGAGAGCGGAAAGTACATCTCTGTGTCTGGGGCTAGCG 694
Db 617 GAAACGCGCAGCGCAGCGCTCCACCGAGCAGACCATACCATTTCTTCGCGAGGGCTGSCA 676
Qy 695 CGGTCAGTTTCTTAGATTTTCAAGAAATCAACTTGCACCAACCGCTTGGACCTGTGTACACA 754
Db 677 GTGATCTCTCTGAGCTTTAAGAAACAGCTCCGCTCTCCGCTGAGCGCTCTGCCACA 736
Qy 755 TTGCGTTTAAACCGGAAGAACGTGCGTTGTACAAAAATATACCGGTTATCTTTAACATCG 814
Db 737 TCAGAAATGACCGCCGAGCGCCAAAGAGTACCGGATCTTCTGTGCTGTTCACGCTG 796
Qy 815 AACGGGGGTTTTCTTTTGAACCCGATGAGGAGCGCGGTGAGATTAAATATGCGATGAAC 874
Db 797 CTAAGGGGTTTCTTCATGGAACCTGATGAGGATAATCATGAGCTTAAGATCTGCGACGAGC 856
Qy 875 ACCGGGCTACACAAATAT-----GGTCCAGAGTGCAGACGCGACCGATGATGAGCATTC 928
Db 857 ATCTTGATATTGCAACTTCTGTCGCGACCCGAGACCGCGGTGAGGTGGCAGTATCC 916
Qy 929 CGTTTCAAAAAAACCCAGATTCCAAAAAGAACCGGCGTTTCGGGGCCCTGCTGAAAG 988
Db 917 CATTTGCAAGCATCATGATCTCTTGAAGCGGAGGCCCGTGCAGAGGACCTTCTCCGTG 976
Qy 989 AGACAATGCCCCAGCTGGCAGACCGTCCATTCAGCTTCGACGCGATTTGCTGTGTGCGCG 1048
Db 977 ATACGATCGCTCATCTTGTGATCGACCACTGTCTTTTGTCTCGTATATGCTGGGATGCTG 1036
Qy 1049 ATACCGGATCGGAATTCCTGATAGATCGACATCCGACATACCAAGTCTTCTGTGTTGG 1108
Db 1037 ATACAGTGGATCGCGCTTCTTGATCGATAGGACATCTCGAGTATCGCTCTTTACTGCTT 1096
Qy 1109 GCTGTGTGCGAGCGGAAGAGGTTTAAATATCTGCTCTTCTATTGGGAATCTCATTTGTTG 1168
Db 1097 CTGTCGTTGATCTGTTAATGAGGCCATGCAATGCTTACCATTTGGTGGTTCATAGCG 1156
Qy 1169 ACAGCATGGAAGTAAAGTCCCGCAAAATATTCAGAAATTAATCAAGTGAACCCGAGACA 1228
Db 1157 ATGCTCTGGAGGAAACCTGCAAAAAGGAACTGAAGCATGCACTACGGTGGAGGCTTGAGA 1216
Qy 1229 TTGGGCGAACCGTAACTGGGCTGATCTCTGGGGCGTTTGGCGGTCCTTAATCGTGA 1288
```

Db 1217 TTGCCGCCCAACGAGACTGGAAGGATACGCAAAATAGATTCCGAGGTCGCAATAAAGTAA 1276

QY 1289 TGGATTTTCATGATGT 1304

Db 1277 TGGATTTCCAAAGGT 1292

RESULT 10

BD173827

LOCUS BD173827 1320 bp DNA linear PAT 18-FEB-2003

DEFINITION Compositions for assaying glycoprotein.

ACCESSION BD173827

VERSION BD173827.1 GI:28415160

KEYWORDS WO 02061119-A/1.

SOURCE unidentified

ORGANISM unidentified

REFERENCE 1 (bases 1 to 1320)

AUTHORS Kozuma,T., Yoshioka,I., Arai,M., Sumitani,J. and Imamu,S.

TITLE Compositions for assaying glycoprotein

JOURNAL Patent: WO 02061119-A 1 08-AUG-2002;

ASAHI KASEI CORP, TAKUJI KOZUMA, ISSEI YOSHIOKA, MOTOO ARAI, JUNICHI SUMITANI, SHIGEYUKI INAMURA

COMMENT OS Fusarium oxysporum IF0-9722

PN WO 02061119-A/1

PD 08-AUG-2002

PF 30-JAN-2002 WO 2002JP000721

PR 31-JAN-2001 JP 01P 022953, 16-FEB-2001 JP 01P 039796 PR

08-AUG-2001 JP 01P 240002

PI TAKUJI KOZUMA, ISSEI YOSHIOKA, MOTOO ARAI, JUNICHI SUMITANI, PI SHIGEYUKI INAMURA

PC C12Q1/26, C12Q1/37, C12N15/09, G01N33/68, G01N33/72 CC

Compositions for assaying glycoprotein

EH Key Location/Qualifiers

FI CDS Location/Qualifiers

1. .1320

/organism="unidentified"

/mol_type="genomic DNA"

/db_xref="taxon:32644"

FEATURES

source

ORIGIN

Query Match 23.6%; Score 335.2; DB 6; Length 1320;

Best Local Similarity 55.3%; Pred. No. 7.2e-88;

Matches 721; Conservative 0; Mismatches 568; Indels 15; Gaps 3;

QY 41 TCACTAAGTCATCATCTCTCTGATGTTGGTGGCGGACTTGGGGACCTCAACGGCTC 100

Db 11 TCACCAACAGTCCCAAAATTCATCGTTGGTGGCGGAACTTGGGGATGCTCAACTGCC 70

QY 101 TGCACCTCGCGCGCGGATATACCAACGTTACCGTGTGGACCCCTATCTGTGCCCTA 160

Db 71 TCCATCTCGCCGTGGGGTTACCAACGTCATCTGTTCTCGATGTCAATCGCATCCGT 130

QY 161 GCGCCATCTCGCGGGAACGAGCTGCAACAAGTCAATTAGCAGTGGCCAATATTCGAATA 220

Db 131 CACCGATATCAGCGGCGCATGATGTAACAACAACTTGCTGCCGACTGTGCGACTGCCGATA 190

QY 221 ACAA-----GACGAATCGAAGTGAATGAGATCTTGGCGGAAGAGCGGTTAAACGGTT 274

Db 191 GCAAAGGTGATGATGAAGACTCAATCTGGAAGACACTTAGCTACGCGCGAGCTCAAGGAT 250

QY 275 GGAAGACGACCCGCTTTTCAACCGTATTATCATGATACGGGCTGCTGATGTCGCTT 334

Db 251 GGCTCACGACCCCTGTTCTTCAACCAATCTGCCACAATACAGGCTCTGTGCTGGTGGCT 310

QY 335 GCTCGGAGGAGGCGCTGGATCGCTGGCGCTCCGGGTACGTCCTCGGGCGGAGGATCCTAATC 394

Db 311 CAACACCAAGTCTATCAACGAGCTGCTAGAGATGAGATCGGTGAGACATCGACCACT 370

QY 395 TGGTGGAACTTACCGGCCCGGAGCAATTCGTAACTGGGCCCGGAAGGGCTGTTGAAG 454

Db 371 ATACACCTCTCAACACAGCAGAAGATTTTCAGAAAGACCATGCGCTGGGGTATCCTGCACG 430

QY 455 GTGATTTTCCGGGTTGAAAAGGTACTTTTCCGGCTTCCGGCGCTGGGCACATGCAA 514

Db 431 GTAACTTTCCAGGCTGGAAGGCTTTTACAAGCCACGGGTTCTGGTGGTTTCATGCTC 490

QY 515 GGAATGCCCTTAGTGGCAGCAGCAGCGCAAGCAGCGCATGGGTGTAATAATTTGTACTG 574

Db 491 GAAAAGCTATGAAAGCTGCTTTTCAAGAGAGAGGAGGCTTGGTGTCAAAATTCATCACTG 550

QY 575 GCACCCCGCAGGTCGTGTAGTCACGTTAATCTTTGAAAATAACGATGTAAGAAGTGCAG 634

Db 551 GCTCTCCGAAAGGAGGTGGAGAGTCTGATCTTTGAAGACGGCGATGTTTCGAGGTGCCA 610

QY 635 TTACGGGCGATGGCAAAATTTTGAGAGCGGAACGTTACATTTCTCTGTGTGGGGCTAGCG 694

Db 611 AGACGGCAGATGTTAAGGAGCAGACAGCGGATCTATCTTTCCGCTGGTGTCTCAG 670

QY 695 CGGGTCAGTTCCTAGATTTCAAGAAATCACTTCGACCAACCGCTTGGACCTCGGTACACA 754

Db 671 CAGAGTTCTTCTCGATTTTGAGAACCAAGTCCAGCTACGGCGTGGACCTCGGGCCATA 730

QY 755 TTGCGTTAAACCGGGAAGACGTCGCTGTGACAAAATATACCGGTTATCTTTAAACATCG 814

Db 731 TCAGATGACACCCAGAAAGAACCAAGCTGTACAAAGAACCTGCCACCTCTTTTCAACAICA 790

QY 815 AACGGGGGTTTTTCTTTGAACCCGATGAGGAGCGCGCTGAGATTAATAATATGCGATGAAC 874

Db 791 ACCAAGTTTCTTCATGGAACCTGTAGAGGATCTTTCATCAACTCAAGATGTGCGATGAAC 850

QY 875 ACCGGGCTACACAATAATGTGTCAGAGTGCAGACGGCAGC---ATGATGAGCATTCCTCGT 931

Db 851 ATCCGGGCTACTGCAACTGGGTTGAAAAGCTGGTTCTAAGTACCCCCAGTCCATCCCT 910

QY 932 TCGAAAAACCCAGATTTCCAAAGAGCGGAAACGCGGTTCCGGCCCTGCTCAAAAGAGA 991

Db 911 TCGAAAGCATCAAGTGCCAACCGAGGCTGAACGACGATGAAGCAGTTTCTGAAAGATA 970

QY 992 CAATGCCCCAGCTGGCAGACCCGTCCATTCAGCTTTCGACGCAATTTGCTGGTGGCCGATA 1051

Db 971 TCATGCTCAGCTTGCAGATCGGCCGCTGTTTCATGCTCGAATCTGCTGGTGGCTGATA 1030

QY 1052 CCGCGAATCGGAATTCCTGTATAGATCGACATCCGAGTACCAAGCTTGTGTGGGCT 1111

Db 1031 CACAGATAGAATGTTCTCTGATCACCTATCATCTCGACATCCCTCACTTGTCAITGCTT 1090

QY 1112 GTGGTCGAGCGGAAGAGGTTTAAATATCTGCTTCTATTGGGAATCTCATTTGTTGACG 1171

Db 1091 CAGGTGATTCGGCACGGGTTACAAGCATATCATCAATTTGAAAGTTTCATCTCTGACT 1150

QY 1172 CGATGGAAGTAAAGTCCGCAAAAAAATTCAGAAATTAATCAAGTGGAAACCCGGACATTG 1231

Db 1151 GTATGAGGTTACGCTTGAGGAAAGGTTGCCAAGTTCTGGAGATGGCGACAGAGAAT 1210

QY 1232 CCGCGAAGCTTAATCTGGCTGATACCTCTGGGGGCTTTTGGCGGTTCAAATCTGTGATGG 1291

Db 1211 TTACCGAGTTCTGGGGTAAAGATCCTCTGGATCGGTTGGAGCTGACGATGAAGATCATGG 1270

QY 1292 ATTTTC-----ATGATGTGAAGAAATGGACCAATGTTCAAGTAT 1329

Db 1271 ATTTGCCAAGAGTGTATGAGGGATGGCAAAATATCAAGAAAT 1314

RESULT 11

E16562

LOCUS E16562 1320 bp DNA linear PAT 28-JUL-1999

DEFINITION Fusarium oxysporum gene for fructosylamine oxidase.

ACCESSION E16562

VERSION E16562.1 GI:5711245

KEYWORDS JP 1998201473-A/1.

SOURCE unidentified

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 1320)

Db	611	AGACGGCAGATGGTAAGGAGCACAGAGCGGATCGAACTATTCTTCCGCTGGTGCTTCAG	670
Qy	695	CGGTCAGTTCCCTAGATTTTCAAGAACTCAACTTCGACCAACCGCTTCGACCCCTGGTACACA	754
Db	671	CAGAGTTCTTCTCGATTTTGAAGACAGATCAGCGCTACGGCGTGGACCTGGGCGCATA	730
Qy	755	TTGCGTTTAAACCGGAGAACGTCGTTGTGTAACAAATATACCGGTTATCTTTTAAACATCG	814
Db	731	TCCAGATGACACACAGAGAAACCAAGCTGTACAAGAACCTGCCACCTCTCTTTTCAACATCA	790
Qy	815	AACGGGGTTTTTCTTTTGAACCCGATGAGGACCGCGTGAGATTAAATATGCGATGAAC	874
Db	791	ACCAAGGTTCTTTCATCGGAACCTGATGAGATCTTCATCAACTCAAGATGTGCGATGAAC	850
Qy	875	ACCGGGCTACACAAATATGGTCAGAGTGCAGACGGCAGC--ATGATGAGCATTTCCGT	931
Db	851	ATCGGGCTACTGCACTGGTTGAAAGCGCTGGTTCTAAGTACCCCGAGTCATCCCT	910
Qy	932	TCGAAAAAACCCAGATTCCAAAAAGAACCGCGCTTCGGGCCCTGCTGTAAGAGA	991
Db	911	TCGCAAGCATCAAGTGCACACCGAGGCTGAACGACGATGAAGCAGTTTCTGAAAGATA	970
Qy	992	CAATGCCCGAGCTGGCAGACCGTCCATTGAGCTTCCGACGCAATTTGCTGGTGGCGGATA	1051
Db	971	TCATGCCCTCAGCTTGCAGATCGGCGCTTGTTCATGCTCGAATCTGCTGGTGGCTGATA	1030
Qy	1052	CCGCGAATCGCAATTCCTGATAGATCGATCATCCGCGATACACAGTCTTGTGTGGGCT	1111
Db	1031	CACAGATGAATGTTCTCTGATCACTATCATCTCGACATCCCTCAGTTGCTTGTCT	1090
Qy	1112	GTGTCGCGCGGAAGAGGGTTTTAAATATCTGCTTCTATTGGGAATCTCATTTGTGAGC	1171
Db	1091	CAGTGATTGCGCACCGGTTTACAAGCATATCACATCAATTCGAAGATTCATCTGACT	1150
Qy	1172	CGATGGAAGTAAATGTCGCGCAAAAAATTCAAGATTAATCAAGTGAACCCGACATTTG	1231
Db	1151	GTATGGAGGGTACGCTTGAGGAAAGGTTTGCCAAATCTCGAGATGGCGACACAGAACT	1210
Qy	1232	CGCGCAACCTGAATCTGCGGTGATACCTCGGGCGTTTTGCGCGTCCAAATCGTGTGATCG	1291
Db	1211	TTACAGAGTTCTGGGGTAAAGATCCCTCTGATCGGTTTGGAGCTGACGATAAGATCATGG	1270
Qy	1292	ATTTC-----ATGATGTGAAGGAATGGACCAATGTTTCAGTAT	1329
Db	1271	ATTGTCCAAGATGATGTAGAGGGATGACAAATATCAAGAT	1314
RESULT 12			
BD173829			
LOCUS		1320 bp	DNA
DEFINITION			linear
ACCESSION			PAT 18-FEB-2003
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

Qy 41 TCACCTAAGTCATCATCTCTCCTGATCGTTGGTGC CGGACCTTGGGGGCACCTCAACGGGTC 100

Db	11	TCACAAACAGTCCCAAATTTCTCATGTGTGGCGGAACTTGGGGATGCTCAACTGCC	70
Qy	101	TGCACCTCGCGCGCGGGATATACCAACGTTTACCGTCTGGGACCCCTCATCTGTGCTCA	160
Db	71	TCCATCTCGCCGTCGGGTTTACACCAACGTCACGTGTTCTCGATGTCAAATCGCATCCCGT	130
Qy	161	GGCGCATCTCGCGCGGAAAACGACGTGAACAAAGTCATTTAGCAGTGGCCAAATATTCGAATA	220
Db	131	CACCGATATACGCGCGGGCATGATGTAAACAAACTTGTGGCGGACTGTGCGATGCCGATA	190
Qy	221	ACAAA-----GACGAAATCGAAGTGAATGAGATCTTGGCGGAAGACGCGTTTAAACGGTT	274
Db	191	GCAAAGGTGATGATGAAGACTCAATCTGGAAGACACTTAGCTACGCCGAGCTCAAGAT	250
Qy	275	GGAAAGACGACCCGCTTTTCAAACCGTATTTATCATGATACGGGCTTGTGATGTCTGCTT	334
Db	251	GGCTCCAGACCCCTGTCTTCCAAACATTTGSCCAATACAGGCTCTGTGCTGGCTGGCT	310
Qy	335	GCTCGCAGGAGGCTTGATCGCTCTGGGCTCGGGTACGTCCGGGCGAGGATCCTTAATC	394
Db	311	CAACACCAAAAGTCTATCAAGCAGCTGTTAGAAGATGAGATCGGTGAACGATCGACCAGT	370
Qy	395	TGTTGGAACTTACC CGCCCGGAGCAATTTCTGTAACTGGCCCGCGGAAGGCGTTTGAAG	454
Db	371	ATACACCTCTCAACACAGCAGAGATTTTCAGAAAGACCATGCCCTGAGGGTATCTCGACAG	430
Qy	455	GTGATTTTCCGGTGTGGAAGGGTACTTTTGGCCGCTTCCGGCGCTGGCTGGGACATGCAA	514
Db	431	GTAATTTTCCAGGCTGGNAGGCTTTTACAGCCACAGGGTTCTGGTTGGTTTCACTGCTC	490
Qy	515	GGAATGCTTAGTGGCAGCAGCACGCGAAGCACAGCGATGGGTGTAAAATTTGTTTACTG	574
Db	491	GAAGAAGCTATGAAGCTGCTTTTCGAAGAGAGCGAGAGGCTTGGTGTCAAATTCATCACTG	550
Qy	575	GCACCCGCGAGGTCGTGTAGTCACGTTAATCTTTGAAATAACGATGTAAAGGTGCGC	634
Db	551	GCTCTCCGGAAGAAAGGTGGAGAGTCTGATCTTTTGAAGACGCGCATGTTTCGAGGTGCCA	610
Qy	635	TTACGGCGCATGGCAAAATTTGGAGAGCGGAACGTACATTCCTGTGTCTGGGGCTAGCG	694
Db	611	AGACGGCAGATGTTAAGGACACAGAGCGATCGAATTTCTTTCCGCTGGTGTCTTCAG	670
Qy	695	CGGTCAGTTCCTAGATTTCAAGAAATCAACTTCGACCAACCGCTTGGACCTTGGTACACA	754
Db	671	CAGAGTTCTTCTCGATTTTGAGAAACAGATCCAGCCTACGGCGTGGACCTGGGCCATA	730
Qy	755	TTGCGTTAAAACCGGAAGACGTGCTGTGTACAAAATAATACCGGTTATCTTTAACTG	814
Db	731	TCCAGATGACACCAAGAAACCAAGCTGTACAGAAACCTGCCACCTCTTTTCAACATCA	790
Qy	815	AACGGGGTTTTTCTTTGAAACCGATGAGGAGCGCGTGAAGTTAAATATGCGATGAAC	874
Db	791	ACCAAGGTTTCTTATGGAACTGATGAGGATCTTTCATCACTCAAGATGTCGATGAC	850
Qy	875	ACCGGGCTACACAAATATGCTCCAGAGTGCAGACGGCACG---ATGATGAGCATTTCCGT	931
Db	851	ATCCGGGCTACTGCAACTGGGGTTGAAAGCTGTGTTCTAAGTACCCCGGTCATCCCT	910
Qy	932	TCGAAAAAACCCAGATTCAAAAGAACGCGAAGACGGCGTTTCGGGCCCTGCTGAAGAGA	991
Db	911	TCGCAAAAGCATCAAGTGCACACCGAGGCTGAACGACGCATGAAGCAGTTCTTGAAGATA	970
Qy	992	CAATGCCCGCAGCTGCAGACCGTCCATTTACGTTTCGACGCGATTTGCTGTGTCGCCGATA	1051
Db	971	TCATGCTCTGACGTTGCGAGATCGGCGGCTTGTTCATGCTCGAATCTGCTGGTGGCTGATA	1030
Qy	1052	CCGCGAATCGGAAATTCCTGATAGATCGACATCCGCGAGTACCAAGTCTTGTGTGGGCT	1111
Db	1031	CACAGGATAGNATGTTTCTGTATCACCCTATCATCTCTGACATCCCTCACTGTGATGCTT	1090
Qy	1112	GTGTCGAGCGGAAGGGTTTAAATATCTGCTCTTCTATTTGGGAATCTCATTTGTCAGC	1171
Db	1091	CAGGTGATTCGGCACGGGTTACTTTCGATATCAATCAATTTGAAGAGTTTCACTCTGACT	1150

QY	1172	CGATGGAAGTAAAGTCCGCCCAAAAATTAATCAAGAAATTAATCAAGTGAACCCGACATTTG	1231
Db	1151	GTATGGAGGGTACGCTTTGAGGAAAGGTTTGCCAAAGTTCTGGAGATGGCCGACCAAGAAGT	1210
QY	1232	CGGCGAACCGTAAATGSCGTGATACTCTGGGGCGTCTTTGGCGGTCCAAATCGTGTGATGG	1291
Db	1211	TTACCGAGTTCTGGGGTAAAGATCCTCTGGAATCGGTTTGGAGCTGACGATTAAGATCATGG	1270
QY	1292	ATTTC-----ATGATGTGAAGGAATGGACCAATGTTTCAGTAT	1329
Db	1271	ATTGTCACCAAGAGTGAATGATAGAGGATGGACAAATATCAAGAAT	1314
RESULT 14			
BD173830			
LOCUS	BD173830	1320 bp	DNA
DEFINITION	BD173830	linear	PAT 18-FEB-2003
ACCESSION	BD173830	Compositions for assaying glycoprotein.	
VERSION	BD173830.1	GI:28415163	
KEYWORDS	WO 02061119-A/4.	synthetic construct	
SOURCE		synthetic construct	
ORGANISM		other sequences; artificial sequences.	
REFERENCE		1 (bases 1 to 1320)	
AUTHORS		Kozuma,T., Yoshioka,I., Arai,M., Sumitani,J. and Imamu,S.	
TITLE		Compositions for assaying glycoprotein	
JOURNAL		Patent: WO 02061119-A 4 08-AUG-2002;	
		ASAHI KASEI CORP, TAKUJI KOZUMA, ISSEI YOSHIOKA, MOTOO ARAI, JUNICHI SUMITANI, SHIGESUKI IMAMURA	
COMMENT		OS Artificial Sequence	
		PN WO 02061119-A/4	
		PD 08-AUG-2002	
		PF 30-JAN-2002 WO 2002JP0000721	
		PR 31-JAN-2001 JP 01P 032953, 16-FEB-2001 JP 01P 039796 PR	
		08-AUG-2001 JP 01P 240002	
		PI TAKUJI KOZUMA, ISSEI YOSHIOKA, MOTOO ARAI, JUNICHI SUMITANI, PI SHIGESUKI IMAMURA	
		PC C12Q1/26, C12Q1/37, C12N15/09, G01N33/68, G01N33/72 CC	
		Compositions for assaying glycoprotein	
		FH Key	
		FT CDS	
		FT mutation	
FEATURES		Location/Qualifiers	
source		1. .1320	
		/organism="synthetic construct"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:32630"	
ORIGIN			

	Query Match	23.4%;	Score 332;	DB 6;	Length 1320;
	Best Local Similarity	55.1%;	Pred. No. 6.4e-87;		
	Matches	719;	Conservative 0;	Mismatches 570;	Indels 15; Gaps 3;
Qy	41	TCACCTAAGTCATCATCTCTCTCTGATCGTTGGTTCGCGGACTTCGGGCACCTCAACGGCTC	100		
Db	11	TCACCAACAGTCCCAAAATTCATCTGTTGGTGGCGAACTTCGGGATGCTCAACTGCC	70		
Qy	101	TGCACCTCGCGCGCGCGGATATACCAACGTTACCGTGTCTGGACCCCTATCTCTGTCCCTA	160		
Db	71	TCCATCTCGCCCGTCGGGGTTACCAACGCTACTGTTCTCGATGTCAATCGCATCCCGT	130		
Qy	161	GCGCCATCTCCGCGGGAACAGCGTGAACAAAGTCATTAGCAGTGGCCAAATATTCGAATA	220		
Db	131	CACGGATATCAGCGGGCATGATGTAAACAACTTGCTGCCGACTGTCGACTGCCGATA	190		
Qy	221	ACAAA-----GACGAATCGAAGTGAATGAGATCTTGGCGGAAGGGCGTTTAAACGGTT	274		
Db	191	GCAAGGTGATGATGAAGACTCAATCTGGAAGACCACTTAGCTACGCCGCGAGCTCAAGGAT	250		
Qy	275	GGAGAACGACCGCGTTTTCAAACCGGTATTATCATGATACGCGGCGTGTGATGTCTGC	334		
Db	251	GGCTCCACGACCTGTCTTCCAAACCACTTCGCCCACAATACAGGCTCTGTCTGGCTGGCT	310		

```
QY 335 GCTCGCAGGAGGCGCTCGGATCGCCTGGCGTCCGGGTACGTCGCGCGGAGGATCCTTAATC 394
Db 311 CAAACCAAAAGTCTATCAAGCAGCTGTGAGAAGATGAGATCGGTGACGACATCGACCAAGT 370
QY 395 TGGTGAACCTTACCCGCGCGGAGCAATTTCTGTAACCTGGCCCGGAAAGCGGTGTTCGAAG 454
Db 371 ATACACCTCTCAACACAGCAGAGAATTTTCAGAAAGACCATGCTGAGGGTATCCTGACAG 430
QY 455 GTGATTTTCGGGTTGGAAGGTAATTTGCGGTTTCGGCGTTCGGCGTGGCGACATGCAA 514
Db 431 GTAACTTTTCAGGCTGGAAGGCTTTTCAAGCCCAAGGCTTCTGTTGGGTTTCATGCTC 490
QY 515 GGAATGCTTAGTGCGCAGCAGCAGCGGAGCAGCAGCGCATGGGTGTAAATTTGTTACTG 574
Db 491 GAAAAGCTATGAAGCTGCTTTTCGAGAGAGCGAGAGGCTTGTGTCAAAATTCATCACTG 550
QY 575 GCACCCGCGAGGTCGTGTAGTCACTTAATCTTTGAAAATAACGATGTAAGAGTGCGG 634
Db 551 GCTCTCCGAAAGGAGGAGTCTGATCTTTGAAGACGCGGATGTTGAGGTGCCA 610
QY 635 TTACGGGCGATGCAAAATTTGAGAGCGGAAGCTATCTCTGTGTGCTGGGCTAGCG 694
Db 611 AGACGGCAGATGTAAGGAGCAGAGCGGATCGAACTATTTCTTCGCTGGTCTCAG 670
QY 695 CGGTCAGTTTCTTAGATTTTCAAGAAATCAACTTCGACCAACCGCTTGGACCCCTGGTACACA 754
Db 671 CAGAGTTCTCTCGATTTTGAGAACAGATCCAGCTACGGCTAGCGGCTGGACCCCTGGGCCATA 730
QY 755 TTGCGTTAAACCGGGAAGACGTCGTTGTGTAACAAATATACCGGTTATCTTTTAAACATCG 814
Db 731 TCCAGATGACACCAGAAGAAACCAAGCTGTACAAAGAACCTGCCACCTCTTTTCAACATCA 790
QY 815 AACGGGGTTTCTTTGTAACCCGATGAGGAGCGCGTGAGATTAAATATGCGATGAC 874
Db 791 ACCAAGGTTTCTTATGGAACCTGTAGGAGATCTTCAATCAACTCAAGATGTGCGATGAAC 850
QY 875 ACCGGGCTACACAAATATGTCAGATGTCAGACGCGACG--ATGATGAGCATTTCCGT 931
Db 851 ATCCGGGCTACTCAACTGCTGGTTGAAAGCTGGTTCTAAGTACCCCGATCCATCCCT 910
QY 932 TCGAAAAAACCCAGATTCCAAAAAGAGCCGAGTTCGCGGCTTCGCGCCCTGCTGAAAGAGA 991
Db 911 TCGAAAGCATCAAGTGCCAAACCGAGGCTGAAACGATGAAGCAGTTTCTGAAAGATA 970
QY 992 CAATGCCCGAGCTGGCAGACCGTCCATTGAGCTTCGACGCCATTTGCTGGTGGCCGATA 1051
Db 971 TCATGCTCAGCTTGCGATCGGCGGCTTGTTCATGCTCGAATCTGCTGGTGGCTGATA 1030
QY 1052 CCGGAATCGGAATTCCTGATAGATCGACATCCGAGTACCACAGTCTTTGTGTGGGCT 1111
Db 1031 CACAGGATAGAAATGTTCTCTGATCACCTATCATCTCTGACATCCCTCACTTGTGATGCTT 1090
QY 1112 GTGGTGGAGCGGAAGGGTTTAAATATCTGCTTTCTATTTGGGAATCTCAATTTTGACG 1171
Db 1091 CAGGTGATTCCGCGACGGTTACGTGCATATCATCAATTTGGAAGTTTCATCTGACT 1150
QY 1172 CGATGGAAGTAAGTCCGCCAAAAAATTCAGAAATTAATCAAGTGAACCCGACATG 1231
Db 1151 GTATGGAAGGTACGCTTGAGGAAGGTTTCCCAAGTTCTGGAGATGGCGACGAGAGT 1210
QY 1232 CGCGAACCGCTAACTGCGGTGATACTCTGGGGGTTTGGCGGTTCGAAATCGTGTATGG 1291
Db 1211 TTACCGAGTTCTGGGTAAAGATCCCTCTGATCGGTTTGAGCTGACGATTAAGATCATGG 1270
QY 1292 ATTTTC-----ATGATGTGAAGAAATGGAACCAATGTTCAGTAT 1329
Db 1271 ATTTGCCAAAGAGTGTAGAGGGATGGACAAATATCAAGAAT 1314
```

RESULT 15
CR382133_02/c
WPCOMMENT

Sequence split into 13 fragments LOCUS CR382133 Accession CR382133

Fragment Name	Begin	End
CR382133_00	1	110000
CR382133_01	100001	210000
CR382133_02	200001	310000
CR382133_03	300001	410000
CR382133_04	400001	510000
CR382133_05	500001	610000
CR382133_06	600001	710000
CR382133_07	700001	810000
CR382133_08	800001	910000
CR382133_09	900001	1010000
CR382133_10	1000001	1110000
CR382133_11	1100001	1210000
CR382133_12	1200001	1249565

Continuation (3 of 13) of CR382133 from base 200001 (CR382133 Debaryomyces hansenii chrom

Query Match 8.9%; Score 127; DB 8; Length 110000;

Best Local Similarity 49.1%; Pred. No. 1.6e-25;

Matches 516; Conservative 0; Mismatches 490; Indels 45; Gaps 5;

QY	62	TGATCGTTGGTCCGCGGACTTGGGGCACCTCAACGGCTCTGCACCTCGCGCGCGCGGAT	121
Db <td>39034</td> <td>TGATTATAGGTGGTGGCACCTGGGGTTTATCAACTGCCCTTTATTTGTTGCGTGATAATT<td>38975</td></td>	39034	TGATTATAGGTGGTGGCACCTGGGGTTTATCAACTGCCCTTTATTTGTTGCGTGATAATT <td>38975</td>	38975
QY <td>122</td> <td>ATACCAACCGTTACCGTGTGACCCCTATCTCTGCTAGCGCCATCTCCGCCGGAACG<td>181</td></td>	122	ATACCAACCGTTACCGTGTGACCCCTATCTCTGCTAGCGCCATCTCCGCCGGAACG <td>181</td>	181
Db <td>38974</td> <td>ATACAGATGTACTATAATCGATGCTGAAGAAATACCATCTTCTAGATCAGCAGGAAATG<td>38915</td></td>	38974	ATACAGATGTACTATAATCGATGCTGAAGAAATACCATCTTCTAGATCAGCAGGAAATG <td>38915</td>	38915
QY <td>182</td> <td>ACGTGAACAAAGTCATTAGCAGTGGCCAATATTTGCAATTAACAAAGACGAAATCGAAGTGA<td>241</td></td>	182	ACGTGAACAAAGTCATTAGCAGTGGCCAATATTTGCAATTAACAAAGACGAAATCGAAGTGA <td>241</td>	241
Db <td>38914</td> <td>ATATAAATAAAATAATTAAGATAGGGTATACCTCT-----GAGTTTC<td>38873</td></td>	38914	ATATAAATAAAATAATTAAGATAGGGTATACCTCT-----GAGTTTC <td>38873</td>	38873
QY <td>242</td> <td>ATCAGATCTTGGCGGAGAGGGCTTTAACGTTTGAAGAACGACCCGCTTTTCAAAACCGT<td>301</td></td>	242	ATCAGATCTTGGCGGAGAGGGCTTTAACGTTTGAAGAACGACCCGCTTTTCAAAACCGT <td>301</td>	301
Db <td>38872</td> <td>ATTACGCTTAGAGTTAGAGGCGAGCAGAAATATGAGAAACGATCTTATTTCAAACCGT<td>38813</td></td>	38872	ATTACGCTTAGAGTTAGAGGCGAGCAGAAATATGAGAAACGATCTTATTTCAAACCGT <td>38813</td>	38813
QY <td>302</td> <td>ATTATCATGATACGGGCTGCTGATGCTGCTTGTCTCGCAGGAGGCGCTGGATCGCCTGG<td>361</td></td>	302	ATTATCATGATACGGGCTGCTGATGCTGCTTGTCTCGCAGGAGGCGCTGGATCGCCTGG <td>361</td>	361
Db <td>38812</td> <td>TTTTTACGAAGTTGGCTATCTAGTCTGTACGCTCTGAAGAGCACCCCTTGGAGAGTTGA<td>38753</td></td>	38812	TTTTTACGAAGTTGGCTATCTAGTCTGTACGCTCTGAAGAGCACCCCTTGGAGAGTTGA <td>38753</td>	38753
QY <td>362</td> <td>GGGTCGGGTGAGTCCGGCGAGGATCCTTA-----ATCTGTGGAACTTACCCGCC<td>412</td></td>	362	GGGTCGGGTGAGTCCGGCGAGGATCCTTA-----ATCTGTGGAACTTACCCGCC <td>412</td>	412
Db <td>38752</td> <td>ATCTCATTAAGAAATCAGATTGGATTAACACAGGAAAGATTGGAGGAACCTTAATACTA<td>38693</td></td>	38752	ATCTCATTAAGAAATCAGATTGGATTAACACAGGAAAGATTGGAGGAACCTTAATACTA <td>38693</td>	38693
QY <td>413</td> <td>CGGAGCAATTTCTGTAACCTGGCCCGGAGGCGTGTGCAAGGTGATTTTCCGGGTTGGA<td>472</td></td>	413	CGGAGCAATTTCTGTAACCTGGCCCGGAGGCGTGTGCAAGGTGATTTTCCGGGTTGGA <td>472</td>	472
Db <td>38692</td> <td>AGGAGGACATCATTTTCAGGCGCCCC---AATTCTTCAAAAGGACATTTACCTGGCTGGC<td>38636</td></td>	38692	AGGAGGACATCATTTTCAGGCGCCCC---AATTCTTCAAAAGGACATTTACCTGGCTGGC <td>38636</td>	38636
QY <td>473</td> <td>AAGGGTACTTTTGGCGCTTCGGCGCTGGGTCGATGCTGCGACATGCAAGAAATGCTTGTGGCAG<td>532</td></td>	473	AAGGGTACTTTTGGCGCTTCGGCGCTGGGTCGATGCTGCGACATGCAAGAAATGCTTGTGGCAG <td>532</td>	532
Db <td>38635</td> <td>GTGGCTATCTACAAGAAATGAAGTGGATGGTGATGCTGCTCCACACAGATGAGAGCAG<td>38576</td></td>	38635	GTGGCTATCTACAAGAAATGAAGTGGATGGTGATGCTGCTCCACACAGATGAGAGCAG <td>38576</td>	38576
QY <td>533</td> <td>CAGACCGGAAGCACACGCGATGGGTGTAAAATTTGTTACTGGCACCCCGCAGGGTCGTG<td>592</td></td>	533	CAGACCGGAAGCACACGCGATGGGTGTAAAATTTGTTACTGGCACCCCGCAGGGTCGTG <td>592</td>	592
Db <td>38575</td> <td>TGGCTAATGAGATTTCCGCGCTTGGGAGGAAAAATTTGTCACT-----GATGAAA<td>38528</td></td>	38575	TGGCTAATGAGATTTCCGCGCTTGGGAGGAAAAATTTGTCACT-----GATGAAA <td>38528</td>	38528
QY <td>593</td> <td>TAGTCAAGTTAATCTTTTGAATAACGATCTAAAGGTGCGGTTACGGCGATGGCAAAA<td>652</td></td>	593	TAGTCAAGTTAATCTTTTGAATAACGATCTAAAGGTGCGGTTACGGCGATGGCAAAA <td>652</td>	652
Db <td>38527</td> <td>TATCGCTTTTATTATTGAGGCTGGAAGTGTGAAGGGGTGCAAGATTTCACCGGAAGG<td>38468</td></td>	38527	TATCGCTTTTATTATTGAGGCTGGAAGTGTGAAGGGGTGCAAGATTTCACCGGAAGG <td>38468</td>	38468
QY <td>653</td> <td>TTTGGAGAGCGGAACCATCTCTGCTGGGCTAGCGGGGTAGCGGGGTCACTTCTTAGATT<td>712</td></td>	653	TTTGGAGAGCGGAACCATCTCTGCTGGGCTAGCGGGGTAGCGGGGTCACTTCTTAGATT <td>712</td>	712
Db <td>38467</td> <td>AGTACTTTGCCGATATAAATCTATCTACTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTT<td>38408</td></td>	38467	AGTACTTTGCCGATATAAATCTATCTACTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTT <td>38408</td>	38408
QY <td>713</td> <td>TCAAGAAATCAACTTCGACCAACCGCTTGGACCTCGGTACACATTCGTTGGTTAAACCGGAAG<td>772</td></td>	713	TCAAGAAATCAACTTCGACCAACCGCTTGGACCTCGGTACACATTCGTTGGTTAAACCGGAAG <td>772</td>	772
Db <td>38407</td> <td>TCAAAAACCAAGCTTATGGCTTAAATGTGGACCTTGGGTGACATTAAGTTAAGTACAGAG<td>38348</td></td>	38407	TCAAAAACCAAGCTTATGGCTTAAATGTGGACCTTGGGTGACATTAAGTTAAGTACAGAG <td>38348</td>	38348
QY <td>773</td> <td>AACGTGGGTGTACAAAATATACCGGTTATCTTTAACATCGAACCGGGGTTTTTCTTTTG<td>832</td></td>	773	AACGTGGGTGTACAAAATATACCGGTTATCTTTAACATCGAACCGGGGTTTTTCTTTTG <td>832</td>	832
Db <td>38347</td> <td>AATCCGAAAGATTCAAAGACATTTCCAGTTATTAGAAGCTTAGAAGGTGGATTTTCTTCTC<td>38288</td></td>	38347	AATCCGAAAGATTCAAAGACATTTCCAGTTATTAGAAGCTTAGAAGGTGGATTTTCTTCTC <td>38288</td>	38288

```
Qy 833 AACCCGATGAGCGCGGTGAGATTAAATATGCGATGAACACCGCGGTACACAATA 892
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
38287 AACCTGATGAAA---CAATGAGATGAAATCTGTAATGAGTTTCCAGGATATACACATT 38231

Qy 893 TGGTCCAGAGTGCAGACGCGCACGATGATGAGCATTCCGTTTCGAAAAAACCCAGATTCCAA 952
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
38230 ATATAAGTGAACAGAGGAACCAATACTTCAGTTCCGGAATATATCAACACTGTTCTCTG 38171

Qy 953 AAGNAGCCGAAACGCGCGTTTCGGGCCCTGCTGAAGAGACAAATGCCCCAGCTGGCAGACC 1012
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
38170 TTGAAGCTGAAAAAGGAATTAGAAAGCTATTGAAAGAAACCTATTCCAGATTTTAGGGATA 38111

Qy 1013 GTCCATTTCAGTTCCGACGCAATTTGCTGGTGTGCGGATACCGGAATCGCGAATTCCTGA 1072
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
38110 AACCGATTGTTATGCTCTAAATTTGTTGGTGTACAGATACCCACAGATATTATTATTA 38051

Qy 1073 TAGATCGACATCCGCGAGTACCACAGTCTTGT 1103
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
38050 TCGACGAATACCCAGAAAGCAATGGTAGTTT 38020
```

Search completed: May 29, 2005, 23:28:26
Job time : 6440 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2005, 19:14:08 ; Search time 837 Seconds
(without alignments)
10035.977 Million cell updates/sec

Title: US-10-622-893A-6
Perfect score: 1419
Sequence: 1 atggagggttcgggtgacga.....atcacatcatcatcattaa 1419

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Geneseqn16Dec04:*
2: Geneseqn1980s:*
3: Geneseqn1990s:*
4: Geneseqn2000s:*
5: Geneseqn2001as:*
6: Geneseqn2001bs:*
7: Geneseqn2002as:*
8: Geneseqn2002bs:*
9: Geneseqn2003as:*
10: Geneseqn2003bs:*
11: Geneseqn2003cs:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	685	48.3	1314	2	AAV09626
2	685	48.3	1314	2	Az07721 Heat-resi
3	683.4	48.2	1314	2	Aav00732 Aspergill
4	335.2	23.6	1320	2	AAV44893 Fructosyl
5	335.2	23.6	1320	6	ABK90475 Fusarium
6	333.6	23.5	1320	6	ABK90477
7	332	23.4	1320	6	ABK90476
8	332	23.4	1320	6	ABK90478
9	324.4	22.9	1335	12	ADM78788
10	115.2	8.1	1284	13	ADSL5772
11	78.6	5.5	1296	10	ADE81212
12	78.6	5.5	72149	10	ADSL1173
13	62.8	4.4	1314	8	ACC48873
14	57.6	4.1	1314	2	AXL15949
15	53.6	3.8	1314	2	AXL15949
16	53.6	3.8	1314	2	AXL15950
17	50.4	3.6	1314	8	ACC48872
18	46.4	3.3	1422	12	ADM78786
19	44.8	3.2	2000	8	ADA71938
20	38.8	2.7	3731	12	ADO39657

21	37.2	2.6	1302	4	AAS56010
22	37	2.6	951	8	ACA37656
23	36.4	2.6	1131	10	ACC61522
24	36.4	2.6	1131	10	ADK64569
25	36.4	2.6	5983	11	ADP65808
26	36.4	2.6	5983	11	ADP65728
27	36.4	2.6	5983	11	ADP65712
28	36.4	2.6	5983	11	ADP65793
29	35.6	2.5	414	9	ACH46271
30	35	2.5	1976	13	ADS56141
31	35	2.5	2000	8	ADA71938
32	35	2.5	2953	6	ABV72307
33	34.8	2.5	222	12	ACH82985
34	34.8	2.5	463	4	ABA57449
35	34.8	2.5	463	4	AAI36997
36	34.8	2.5	463	4	AAK31088
37	34.8	2.5	463	4	AAK05487
38	34.8	2.5	463	4	ABS30767
39	34.8	2.5	463	6	ABS05839
40	34.8	2.5	593	12	ACH89285
41	34.6	2.4	456	5	AAH66227
42	34.6	2.4	456	8	ACA00431
43	34.6	2.4	475	11	ADL65880
44	34.6	2.4	5059	2	AAH84332
45	34.6	2.4	349980	5	AAH68527

ALIGNMENTS

RESULT 1

AAV09626
ID AAV09626 standard; cDNA to mRNA; 1314 BP.
XX
AC AAV09626;
XX
DT 14-MAY-1998 (first entry)
XX
DE A. terreus FAOD-L cDNA.
KW Fructosylamino acid oxidase; FAOD-L; Amadori compound; ds.
XX Aspergillus terreus.
XX JPI0033180-A.
XX
PD 10-FEB-1998.
XX
PF 24-JUL-1996; 96JP-00194557.
XX
PR 24-JUL-1996; 96JP-00194557.
XX
PA (KYOT-) KYOTO DAIICHI KAGAKU KK.
XX
DR WPI; 1998-172097/16.
DR P-PSDB; AAW39253.
XX
PT Aspergillus terreus recombinant fructosylamino acid oxidase - used for analysis of "Amadori" compound(s).
XX
PS Example 1; Col 23-26; 18pp; Japanese.

CC This sequence encodes a novel recombinant fructosylamino acid oxidase (FAOD-L) produced from eukaryotic cells integrated with an RT-PCR product encoding FAOD-L derived from Aspergillus terreus Gp1 (FERM P-15664). FAOD-L can be used for the analysis of "Amadori" compounds

SQ Sequence 1314 BP; 294 A; 371 C; 395 G; 254 T; 0 U; 0 Other;

Query Match 48.3%; Score 685; DB 2; Length 1314;
Best Local Similarity 70.7%; Pred. No. 2.5e-218;
Matches 926; Conservative 0; Mismatches 380; Indels 3; Gaps 1;

Aas56010 Salmonell
Aac37656 Prokaryot
Acc61522 Gene seque
Adk64569 Disease t
Adp65808 Human mRN
Adp65728 Human KIA
Adp65712 Human KIA
Adp65793 Human mRN
Ach46271 Human inf
Ads56141 Bacterial
Ada71938 Rice gene
Abv72307 Nucleotid
Ach82985 Human gen
Aba57449 Human toe
Aai36997 Probe #56
Aak31088 Human bon
Aak05487 Human bra
Abe30767 Human liv
Abe05839 Human gen
Ach89285 Human gen
Aah66227 C glutami
Aca00431 C. glutam
Adl65880 C. glutam
Aax84332 Stealth v
Aah68527 C glutami

QY	38	CCGTCACTAAGTCATCATCTCTCTGATCGTTGGTCCGGGACTTGGGGCACCTCAACGG	97
Db	5	CAGTCACCAAGTCTTTCGTGATATTTGATCATCGGGCGGGCACCTGGGGTGTCTCAACTG	64
QY	98	CTCTGACCTCGCGCGCGGAGATATACCAAGTTACCGTGTGGACCCCTATCTGTGCC	157
Db	65	CCCTGTCATCTTGGCCGAGAGGATACACCAATGTCACTGTCTTGACCCCGTACCCGGTTC	124
QY	158	CTAGCGCATCTCCGCGGAAACGAGTGAACAAAGTCATTAGCAGTGGCCATATTCGA	217
Db	125	CATCAGCCATTTCGGCGGCAACGAGTCAACAGATCATCTCGTCGGCCAGTACAGCA	184
QY	218	ATAACAAAGACGAATCGAAGTGAATGAGATCTTTGGCGGAAGGGGTTTTAAAGTTGGA	277
Db	185	GCAAGAAGGACGAGTTCGAAGTCAATGAGATTATCGCCGAACAGGCTTCAATGGCTGA	244
QY	278	AGAACGACCCGCTTTTCAAAACCGTATATCATGATACGGGCCCTGTGTGCTGTGCT	337
Db	245	AAATGACCCCATCTTCAAGCCGTACTACCACGACACCGCGCTGTGATGTCCGCCACCA	304
QY	338	CGCAGGAGGSCCTGGATCGCTCGGCGTCCGGGTACGTCCGGCGGAGGATCCTAACTGG	397
Db	305	CACAGGAAGATTGGAGCGTCTGGGGTCCGGTGGACCTGAAGATGAACCCGATGTAG	364
QY	398	TGGAACCTTACCCCGCGGACAAATTCGTAAACTTGGCCCCGGAAGCGTGTTCGAAAGTG	457
Db	365	CCGAATTGACTCGGCCGAGCAGTTCGCCCAGCTGGCCCC---GGGCTCTGAAGGTA	421
QY	458	ATTTTCGGGTTGGAAGGTACTTTTGGCGGTTTCGGCGCTGTGGGACACATGCAAGGA	517
Db	422	ACTTCCCGGTTGGAGGGGTACCAATTCGCTCAAAACGCGGCTGGCGCATGCGCGCA	481
QY	518	ATGCCTTAGTGGCAGCAGCAGCGAGCAGCAGCGCATGGGTGAATAATTTGTTACTGGCA	577
Db	482	ACGCCCTTGGTCGCGCGCGCGGAGGACACAGCGCTGGGTGTGCGCTTCGTCCGCGGAT	541
QY	578	CCCGCAGGTCGTGTAGTCACGTTAATCTTTGAAATAACGATGTAAAGGTGCGCGTTA	637
Db	542	CGCGCAGGCGCAGATCATCACTTGTATTTTGAACAACGATGTCAGGGTGCCTCA	601
QY	638	CGGCGATGCGCAAAATTTGAGAGCGGAACGTACATTCCTGTGTGTGGGGCTAGCGCGG	697
Db	602	CGCGCGCAGCAAGATCTGGCGGCGCAGCAGCATATCTCTCGCTGTGTGGCGCGCG	661
QY	698	GTCAAGTCTTAGATTTCAAGATCACTTCGACCAACCGCTTCGACCTGGGTACACATTG	757
Db	662	GCCAGTTCTTGGATTTCAAGGACCACTGCTGCCCTGGGTGTGCGCTTCGTCCGCGAT	721
QY	758	CGTTAAACCGGAAGACGTGCGTTGTACAAATAATACCGGTTATCTTTAAACATCGAAC	817
Db	722	AGTTGAAGCCGGAAGAGCGTGCCCAAGTATATAAACATGCCGGTGGTCTTCAACATCGA	781
QY	818	GGGGGTTTTCTTTGAACCCGATGAGGAGCGCGGTGAGATTAAATATGCGATGAACACC	877
Db	782	AGGGGTTCTTCTTCGAGCCGATGAGGCGGTGTGAAATCAAGATCTGCGACGAAACACC	841
QY	878	CGGCTACACAAATATGTTCCAGATGTCAGACGCGCAGATGATGAGCATTCGCTTCGAA	937
Db	842	CCGGGTACACGAATATGACCACGGGGGCCGACGCGCGGTGAGGAGATTTCCCTTCGAG	901
QY	938	AAACCCAGATTCCAAAGAGCCGAAACGCGCGTTCGGGGCCCTGCTGAAAGAGACAATGC	997
Db	902	AGACGCGAGTTCTTCGAGAGCGGAGATGCGCTCCGCAAGCTTCTGTCTGAACGATGC	961
QY	998	CCAGCTGGCAGACCGTTCATATGCTTCGAGCAGCATTTGCTGGTGTGCCGATACCGCGA	1057
Db	962	CTCAGCTTGGCGACCGCGCTTCAGTTTTCGCAAGGATCTCTGTGTGTGGGATACCCCA	1021
QY	1058	ATCGCAATTCCTGATAGATCGACATCCGAGTACACAGTCTTGTGTGGGCTGTGGTG	1117
Db	1022	ATCGCGATTTATCATTTGACCGTCACTCCGAAATACCCGTCGTGTGTTCCTGGGTGGTG	1081
QY	1118	CGAGCGAAGAGGGTTTTAAATATATCTGCTTTCTATTGGGAATCTCATTTGTTGACGCGATGG	1177

Db	1082	CTTCAGGACGAGGCTTCAAATATCTTCCCTCGATCGGAAGCATCATCGCAGACGCCATGG	1141
QY	1178	AAGTTAAAGTGGCGCAAAAAAATTCAGAAATTAATCAAGTGGAAACCGGACATTTGCGGCGGA	1237
Db	1142	AGGACAAAACCCCGGCAAAAATCCACAAGCTGATCGCTGGAGCCCGGAATCGCGATCA	1201
QY	1238	ACCGTAACCTGGCGTGATCTCTGGGGCGTTTTTGGCGGTCCAAAATCGTGTGATCGATTTTC	1297
Db	1202	ACCGTAACCTGGGGGACAGATTAGTTCGATTTCGAGGGGCCCAACCGGTCATGGATTTCA	1261
QY	1298	ATGATGTGAAGGAATGAGCAATGTTTCAGTATCGTGTGATTTTCCAAGCT	1346
Db	1262	ATGAAGTGAAGGAGTGAGTCTAATGTCAACCAAGGACATCTCGAAGTT	1310
RESULT 2			
AAZ07721			
ID	AAZ07721	standard; cDNA to mRNA; 1314 BP.	
XX	AAZ07721;		
AC	AAZ07721;		
XX	02-NOV-1999	(first entry)	
DT	02-NOV-1999	(first entry)	
XX		Heat-resistant fructosyl amino acid oxidase nucleotide sequence.	
DE		PAOD; thermostable; fructosyl amino acid oxidase; heat-resistant;	
XX		glutathione S-transferase; GST; Amadori compound; ds.	
KW			
KW			
XX		Aspergillus terreus.	
OS			
XX		Location/Qualifiers	
Key		1..1314	
PH		/*tag= a	
FT		/product= "fructosyl amino acid oxidase (PAOD)"	
FT			
XX	JP11221081-A.		
PN	17-AUG-1999.		
XX			
PD			
XX			
XX	10-FEB-1998;	98JP-00028169.	
PF	10-FEB-1998;	98JP-00028169.	
XX		(KYOT-) KYOTO DAIICHI KAGAKU KK.	
PA			
XX	WPI; 1999-520719/44.		
DR	P-PSDB; AAY29100.		
DR			
XX		New heat-resistant fructosyl amino acid oxidase - useful for analysis of	
PT		Amadori compounds.	
XX			
PS	Claim 1; Page 5-6; 8pp; Japanese.		
XX			
CC	This nucleotide sequence encodes an Aspergillus terreus heat-resistant		
CC	(thermostable) fructosyl amino acid oxidase (PAOD). PAOD can be produced		
CC	using standard recombinant methodology. A recombinant fusion protein		
CC	comprising PAOD and glutathione S-transferase (GST) is useful for the		
CC	analysis of Amadori compounds		
XX			
SQ	Sequence 1314 BP; 294 A; 371 C; 395 G; 254 T; 0 U; 0 Other;		
Query Match 48.3%; Score 685; DB 2; Length 1314;			
Best Local Similarity 70.7%; Pred. No. 2.5e-218;			
Matches 926; Conservative 0; Mismatches 380; Indels 3; Gaps 1;			
QY	38	CCGTCACTAAGTCATCATCTCTCTGATCGTTGGTCCGGGACTTGGGGCACCTCAACGG	97
Db	5	CAGTCACCAAGTCTTCGTGATATTGATCATCGGGCGGGCACCTGGGGTGTGCTCAACTG	64
QY	98	CTCTGACCTCGCGCGCGGATATACCAAGTTACCGTGTGGACCCCTATCTGTGCC	157
Db	65	CCCTGTCATCTTGGCCGAGAGGATACACCAATGTCACTGTCTTGACCCCGTACCCGGTTC	124

[illegible]

D	b		1142	AGACAAAAAACC CGGCAAAAATCCACAAGTGTATCGCTGGAGCCCCGGAAATCCGCATCA	1200
Q	y		1238	ACCGTAAC TGGCGTGATACTCTGGGCGGTTTTGGCGGTCCAATTCGTGTGATGAATTTC	1297
D	b		1202	ACCGTAAC TGGGGGACAGATTAGGTCGATTTGGAGSGGCCAACCGGGTCATGGA TTCA	1261
Q	y		1298	ATGATGTGAAGGAATGGACCAATGTT CAGTATC GTGATATTTCCAAAGCT	1346
D	b		1262	ATGAAGTGAAGGAGTGGACTAATGTCACCC AAGGACATCTCGAAGTT	1310
 RESULT 4 AAV44893 standard; DNA; 1320 BP.					
X	X	A	AAV44893;		
X	X	D	T	28-OCT-1998 (first entry)	
X	X	D	E	Fructosylamine oxidase coding sequence.	
X	X	K	M	Fructosylamine oxidase; enzyme production; ds.	
X	X	O	S	Fusarium oxysporum.	
X	X	P	N	JF10201473-A.	
X	X	P	D	04-AUG-1998.	
X	X	F	F	20-JAN-1997; 97JP-00007101.	
X	X	P	R	20-JAN-1997; 97JP-00007101.	
X	X	P	A	(ASAH) ASahi KASEI KOgyo KK.	
X	X	D	R	WPI; 1998-474488/41.	
X	X	D	R	p-PsDB; AA W69251.	
X	P	T		Substantially pure microorganism - used for producing fructosylamine	
X	P	T		oxidase efficiently.	
X	P	S		Claim 1; Page 11-13; 15pp; Japanese.	
C	C	C	C	This sequence encodes the fructosylamine oxidase produced by the microbe	
C	C	C	C	of the invention. The microbe can produce fructosylamine oxidase with	
C	C	C	C	high efficiency	
X	S	Q		Sequence 1320 BP; 348 A; 333 C; 330 G; 309 T; 0 U; 0 Other;	
Q	y		41	TC ACTA AG TC AT CAT CTCTCTCTGATCGTTGGTGC GG GACTTGGG CACCTCAACGG CTC	100
D	b		11	TC ACC AAA CA GT CCC AAA TTCTCATCGTTGGTGGCG GA CTGGGGATGCTCAACTGCC C	70
Q	y		101	TG ACCTCGCGCGCGGGATATACCAAAGTTCACGTGCTGGACCCCTATCTCTGTCCCTA	160
D	b		71	TCCATCTCGCGCGGTGGGGGTATACCAACAGTCACTGTTC TGATGTCAATCGCATCCCGT	130
Q	y		161	GCGGCATCTCGCGCGGAAAACGAGCTGAACAAAGTCATTAGCAGTGGGCCAATATTTCGAATA	220
D	b		131	CACCGATATCAGCCGGGCATGATGTAAACAACTTGTGSCCGACTGTCGACTGCCGATA	190
Q	y		221	ACAAA-----GACGAAATCGAAGTAGAATCTTTGGCGGAAGAGCGCTTTAAACGGTT	274
D	b		191	GC AAA GGTGATGATGAAGACTCAATCTGGAAGCACTTAGCTACGCCGCAGCTCAAGGAT	250
Q	y		275	GGAAGAAGCACCGCTTTTCCAAACCGPATTAATCATGATACGGGCTGCTGATGTCTGCTT	334
D	b		251	GGCTCCACGACCCCTGTCTTCAACCACTTTCGCCAATAACAGGCTCTGTCTGTGGTGGCT	310

Qy	335	GCTCGCAGGAGGCGCTGGATCGCTCGGGCGTCCGGGTACGTCGCGGCAGGATCTCTAATC	394
Db	311	CAACACCAAAAGTCTATCAAGCAGCTGGTAGAAGATGAGATCGGTGACGACATGACCCAGT	370
Qy	395	TGTTGGAACTTATCCCGCCCGGAGCAATTTTCGTAACCTGGCCCGCGAAGGCGGTGTGCAAG	454
Db	371	ATACACCTCTCAACACAGCAGAAGATTTTCAGAAAGACCATGCCCTGAGGGTATCTCTGACAG	430
Qy	455	GTGATTTTTCGGGTGTGAAAGGGTACTTTTGGCGGTTTCGGGCGCTGGGTGGGCACATGCAA	514
Db	431	GTAACCTTTCCAGGCTGGAAGGCGCTTTTACAAGCCACAGGGTCTCTGGTTGGGTTCATGCTC	490
Qy	515	GGAATGCTTAGTGGCAGCAGCACGCGNAGCACAGCATGGGTGTAATAATTTGTTTACTG	574
Db	491	GAAGAAGCTATGAAGAGCTGCTTTTCSAAGAGACGAGAGGCTTTGGTGTCAAAATTTCAATCA	550
Qy	575	GCACCCCGCAGGGTCGTGTAGTCACTGTTAATCTTTTGAATAATAACGATGTAAAAGTGCCG	634
Db	551	GCTCTCCGAGGAAAGGTGGAGAGTCTGATCTTTTGAAGACGGCGATGTTTCGAGTGCCTCA	610
Qy	635	TTACGGGCGGATGGCAAAATTTTGGAGAGCGGAACTGATTCCTGTGTCTGGGGCTAGCG	694
Db	611	AGACGGCAGATGGTAAGGAGCACAGAGCGGATCGAACTATTTCTTTCGCTGGTCTTCTCAG	670
Qy	695	CGGTCAGTTCCTAGATTTCAAGAAATCAACTTTTCGACCAACCGCTTGGACCCTGGTACACA	754
Db	671	CAGAGTCTTCTCTCGATTTTGAGAAACAGATCCAGCTTACGGCGTGGACCCCTGGGCCATA	730
Qy	755	TTGCGTTTAAACCGGAAGAACGTCGCTTGTGTAACAAAAATATACCGGTTATCTTTAAACATCG	814
Db	731	TCCAGATGACACACAGAAGAACCAAGCTGTACAAGAACCTGSCACCTCTTTTCAACATCA	790
Qy	815	AACGGGGGTTTTTCTTTGAACCCGATCAGAGAGCGGCTGAGATTTAAATATTCGATGAAC	874
Db	791	ACCAAGGTTTTCTCATGGAACCTGATGAGGATCTTTCATCACTCAAGATGTGCGATGAAC	850
Qy	875	ACCGGGCTACACAAAATATGTCGCAGAGTGCAGACGGCACG- --ATGATGAGCAATTCGGT	931
Db	851	ATCCGGGCTACTGCACCTGGGTTGAAAGCTGGTTCTAAGTACCCCGAGTCCATCCCTCT	910
Qy	932	TCGAAAAAACCCAGATTTCCAAAAGAACCGCAACCGCGGTTTCGGGGCCCTGCTGAAAGAGA	991
Db	911	TCGCAAAAGCATCAAGTGCACACCGAGCTGAACCGACGCATGAAGCAGTTCCTGAAAGATA	970
Qy	992	CAATGCCCGCAGCTGGCAGACCGTCCATTCAGCTTCGCACGCAATTTGCTGGTGGCCGATA	1051
Db	971	TCATGCTCTCAGCTTCGAGATCGGGCGCTGTTTCTATGCTCGAATCTGCTGGTGGCTGATA	1030
Qy	1052	CCGCGAATCGGCAATTCCTGATAGATCGACATCCGCGAGTACCAAGTCTTGTGTGGGCT	1111
Db	1031	CACAGGATAGAAATGTTTCTGATCACTTATCATCTCGACATCCCTCACTGTGCAATGCTT	1090
Qy	1112	GTGTGCGAGCGGAAGAGGGTTTAAATATCTGCCTTCTATTGGGAATCTCATTTGTCAGC	1171
Db	1091	CAGGTGATTCGGCACGGTTTACAAGCATATCACATCAATTTGGAAGATTTCATCTCTGACT	1150
Qy	1172	CGATGGAAGGTAAAGTCGCGCAAAAATTCACGAATTAATCAAGTGGNAACCGGACATTCG	1231
Db	1151	GTATGGAGGGTACCGCTTGAGGAAAGGTTTGCCAAGTTCTTGAGATGGCGACACAGAGAAT	1210
Qy	1232	CGGCGAAACCGTAACTGGCGGTGATACTCTGGGGCGCTTTTGGCGGTCCAAATCGTGTGATGG	1291
Db	1211	TTACCGAGTTCGGGTAAAGATCTCTGGATCGGTTTGGAGTGGCGACGATAAGATCATGG	1270
Qy	1292	ATTTTC-----ATGATGTGAAGAAATGGACCAATGTTTCAGTAT	1329
Db	1271	ATTTGGCCCAAGAGTGTAGTGAGGGATGGGACAAATATCAAGAAT	1314

RESULT 5
ABK90475
ID ABK
XX

AC	ABK90475;
XX	
XX	05-NOV-2002 (first entry)
XX	
DE	Fusarium oxysporum polynucleotide #1.
XX	
KW	Glycino acid; glycoprotein; deoxycholic acid; deoxycholic amide; gene;
KW	cholic acid amide octyl glucoside; quaternary ammonium salt; de;
KW	quaternary ammonium salt cationic surfactant; concanavalin A; betaine;
KW	ascorbic oxidase; 4-(2-hydroxyethyl)-1-piperazinyl group; albumin;
KW	protease; glycoalbumin; bromocresol purple; glycation; diabetes;
KW	ascorbic acid; mutant.
XX	
OS	Fusarium oxysporum.
OS	Synthetic.
XX	
XX	
FH	Key Location/Qualifiers
FT	CDS 1..1320
FT	/*tag= a
FT	/product= "Fusarium oxysporum polypeptide #1"
FT	/partial
FT	/notes= "No start or stop codon given"
FT	/transl_except= (pos:736..738, aa:Ile)
FT	/transl_except= (pos:1186..1188, aa:Iyr)
XX	
PN	W0200261119-A1.
XX	
PD	08-AUG-2002.
XX	
PF	30-JAN-2002; 2002WO-JP000721.
XX	
PR	31-JAN-2001; 2001JP-00022953.
PR	16-FEB-2001; 2001JP-00039796.
PR	08-AUG-2001; 2001JP-00240002.
XX	
FA	(ASAH) ASahi Kasei Kogyo KK.
XX	
PI	Kouzuma T, Yoshioka I, Arai M, Sumitani J, Imamura S;
XX	
DR	WPI; 2002-599854/64.
DR	P-PSDB; ABG30782.
XX	
PT	Compositions for assaying glycoprotein without interference from globulin
PT	and ascorbic acid in blood, useful in clinical examination including
PT	diagnosis and management of diabetes.
XX	
FS	Claim 12; Page 63-65; 82pp; Japanese.
XX	
CC	The invention relates to compositions for assaying glycoprotein with use
CC	of protease and enzymes acting on glycoamino acids in the presence of at
CC	least 1 of deoxycholic acid or amide, cholic acid amide octyl glucoside,
CC	quaternary ammonium salt, quaternary ammonium salt cationic surfactant,
CC	concanavalin A and betaine and/or ascorbic oxidase and a buffer without a
CC	4-(2-hydroxyethyl)-1-piperazinyl group. The invention also relates to
CC	compositions for assaying albumin containing the protease and enzymes for
CC	determining glycoalbumin and a protein-denaturing agent and/or a compound
CC	having S-S bond and bromocresol purple for evaluation of glycation
CC	proportion of albumin into glycoalbumin with respect to the other
CC	albumins measured separately. The compositions are useful in clinical
CC	examination including diagnosis and management of diabetes. The
CC	compositions can be used in methods which avoid the effects of globulin
CC	and ascorbic acid components and stabilise enzymes such as protease
CC	acting on glycoamino acids. This sequence represents DNA encoding a
CC	polypeptide used in the scope of the invention
XX	
SQ	Sequence 1320 BP; 348 A; 333 C; 330 G; 309 T; 0 U; 0 Other;
	Query Match 23.6%; Score 335.2; DB 6; Length 1320;
	Best Local Similarity 55.3%; Pred. No. 7e-101;
	Matches 72; Conservative 0; Mismatches 568; Indels 15; Gaps 3;
QY	41 TCACTAAGTCATCATCTCTCCTGATCGTTGGTGCGGACCTTGGGCACCTCAACGGCTC 100

Db 11 TCACCAACAGTCCCAAAATTCATCGTTGGTGGCGAACTTGGGGATGCTCAACTGCC 70
QY 101 TGCACCTCGCGCCGCGGATATACCAAGTTACCGTGTGACACCGCTATCTGTCCCTA 160
Db 71 TCCATCTCGCGCTGGGGTTACACCAAGTCACTGTTCTCGATGTCATCGATCCCGT 130
QY 161 GGGCCATCTCCGCGGAACGAGCTGAAACAAAGTCAATTAGCAGTGGCCCAATATTCGAATA 220
Db 131 CACCGATATCAGCGGGCATGATGTAAACAAACTTGTGCGCGACTGTGCGACTGCCGATA 190
QY 221 ACAA-----GACGAAATCGAAGTCAATGAGATCTTGGCGGAGAGGCGTTTACGGTT 274
Db 191 GCAAAGGTGATGATGAAGACTCAATCTGGAAGCACTTACGTACGCCGAGCTCAAGGAT 250
QY 275 GGAAGAACACCGCTTTTCAACCGCTATTATCATGATACGGGCTGTGATGCTGTGCTT 334
Db 251 GGCTCACACACCTGTCTTCCACCAATCTGCCCAATACAGGCTCTGTGCGTGGCTGGCT 310
QY 335 GCTCGCAGGAGGCGCTTGGATCGCTCGCGCTCCGGTACGTCCGGCGAGGATCTCTTAATC 394
Db 311 CAAACCAAAAGTCTATCAAGCAGCTGTTAGAGATCGGTGACGACATCGACCACT 370
QY 395 TGGTGAACCTTACCCCGCGGACCAATTTCTTAACTGGCCCCCGAAGGCGTTTCAAG 454
Db 371 ATACACCTCTCAACACAGCAGAGAGATTTAGAAAGACCAATGCTGAGGATCTCTGACAG 430
QY 455 GTGATTTCCGGGTGGAAGGGTACTTTGCGGTTCCGCGCTGCTGGCGACATGCAAA 514
Db 431 GTAACTTTCCAGGCTGGAAGGGCTTTACAGGCCACGGGTTCTGTTGGTTTCATGCTC 490
QY 515 GGAATCCCTTAGTGGCAGCAGCAGCGAGCAGCGCATGGGTGTAATAATTTGTTACTG 574
Db 491 GAAAAGCTATGAAAGCTGCTTTTGAAGAGAGCAGAGGCTTGGTGTCAAATTCATCACTG 550
QY 575 GCACCCGCGAGGTCGTGTAGTACAGTTTAACTTTGAAATACGATGTAAAGGTGCCG 634
Db 551 GCTCTCCGGAAGAAAGGTGGAGTCTGATCTTTGAAGACGCGGATGTCGAGGTGCCA 610
QY 635 TTACGGCGCATGCAAAATTTGAGAGCGGAAGCTACATTTCTGTGTGCGGGCTAGCG 694
Db 611 AGACGCGAGATGTAGGAGCAGACAGCGGATCGAACTATCTTTCGCTGGTGGCTTCAG 670
QY 695 CCGGTTCAGTTCTTAGATTTCAAGATCAACTTCGACCAACCGCTTGGACCCCTGGTACACA 754
Db 671 CAGAGTTCTCTCGATTTTGAGAACAGATCCAGCTACCGCTGCGGCGTGGACCCCTGGCCATA 730
QY 755 TTGCGTTAAACCGGAGAGACGTGGTGTGTAACAAATATACCGTTATCTTTAAATCG 814
Db 731 TCCAGATGACACAGAGAAGAACCAAGCTGTACAGAACCTTGCCACCTCTTTTCAACATCA 790
QY 815 AACGGGGTTTTTTTGTGAACCGGATGAGGCGCGTGAGATTAATAATATGCGATGAAC 874
Db 791 ACCAAGTTTCTTCATGGAACCTGATGAGGATCTTCACTCACTCAAGATGCGGATGAAC 850
QY 875 ACCCGGGCTACACAAATATGGTCCAGATGTCAGACCGGACG---ATGATGAGCATTCGGT 931
Db 851 ATCCGGCTACTGCAACTGGGTTGAAAGCGCTGTTCTAAGTACCCCGCTCCATCCCT 910
QY 932 TCGAAAAACCCAGATTCGAAAGAGCGAAGCGCGCTTCGGGCCCTGCTCAAGAGA 991
Db 911 TCGCAAGCATCAAGTGCCAACCGGCTGAAACGACATGAAGCAGTTTCTGAAAGATA 970
QY 992 CAATGCCAGCTGGCAGACCGTTCATTCAGTTTCGACGCACTTGTGCTGTGCGCGATA 1051
Db 971 TCATGCTCAGCTTGAGATCGGCGCTTGTTCATGCTCGAATCTGCTGTGCTGCTGATA 1030
QY 1052 CCGCGAATCCGGAATTCCTGATAGATGACATCCGCAAGTACCAAGCTTTGTGTGGGCT 1111
Db 1031 CACAGGATAGAAATGTTCTCTCATCACCTATCATCTCGACATCCCTCACTTGTCAATGCTT 1090
QY 1112 GTGGTGGACCGAAGGGTTTAAATATCTGCTTCTTATGGGAATCTCATTTGTGACG 1171
Db 1091 CAGGTGATTGCGGACCGGTTTACAAGCATATCATCAATTTGAAAAGTTCTCTCTGACT 1150

QY 1172 CGATGGAAGGTAAAGTGGCGCAAAAAATTCAGAAATTAATCAAGTGGAAACCGGACATTG 1231
Db 1151 GTATGGAGGTACGCTTGAGGAAAGGTTTTGCCAAGTTCTGGAGATGGCGACAGAGAAGT 1210
QY 1232 CGCGCAACCGTAACTGGCGTGATCTCTGGGCGGTTTGGCGGTCCAAATCGTGTGATGG 1291
Db 1211 TTACCGAGTTCTGGGTTAAAGATCTCTTGGATCGGTTTGGAGCTGACGATAAGATCATGG 1270
QY 1292 ATTTTC-----ATGATGTGAAGGAATGACCAATGTTTCAGTAT 1329
Db 1271 ATTGCCCAGAGTGTATGTAGGGATGGCAAAATATCAAGAT 1314

RESULT 6

ABK90477
ID ABK90477 standard; DNA; 1320 BP.
XX
AC ABK90477;
XX
DT 05-NOV-2002 (first entry)
XX
DE Fusarium oxysporum polynucleotide #3.
XX
KW Glycoamino acid; glycoprotein; deoxycholic acid; deoxycholic amide; gene;
KW cholic acid amide octyl glucoside; quaternary ammonium salt; ds;
KW quaternary ammonium salt cationic surfactant; concanavalin A; betaine;
KW ascorbic oxidase; 4-(2-hydroxyethyl)-1-piperazinyl group; albumin;
KW protease; glycoalbumin; bromocresol purple; glycation; diabetes;
KW ascorbic acid; mutant.
XX
OS Fusarium oxysporum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1320
FT /tag= a
FT /product= "Fusarium oxysporum polypeptide #3"
FT /partial
FT /note= "No start or stop codon given"
FT /transl_except= (pos:736..738, aa:ile)
FT /transl_except= (pos:1186..1188, aa:Tyr)
FT mutation replace(1115,A)
FT /tag= b
XX
PN WO200261119-A1.
XX
PD 08-AUG-2002.
XX
PF 30-JAN-2002; 2002WO-JP000721.
XX
PR 31-JAN-2001; 2001JP-00022953.
PR 16-FEB-2001; 2001JP-00039796.
PR 08-AUG-2001; 2001JP-00240002.
XX
PA (ASAH) ASAH KASEI KOGYO KK.
XX
PI Kouzuma T, Yoshioka I, Arai M, Sumitani J, Imamura S;
XX
XX WPI; 2002-599854/64.
DR P-PSDB; ABG30784.
XX
XX Compositions for assaying glycoprotein without interference from globulin
PT and ascorbic acid in blood, useful in clinical examination including
PT diagnosis and management of diabetes.
XX
PS Disclosure; Page 69-72; 82pp; Japanese.
XX
CC The invention relates to compositions for assaying glycoprotein with use
CC of protease and enzymes acting on glycoamino acids in the presence of at
CC least 1 of deoxycholic acid or amide, cholic acid amide octyl glucoside,
CC quaternary ammonium salt, quaternary ammonium salt cationic surfactant,
CC concanavalin A and betaine and/or ascorbic oxidase and a buffer without a

CC 4-(2-hydroxyethyl)-1-piperazinyl group. The invention also relates to
CC compositions for assaying albumin containing the protease and enzymes for
CC determining glycoalbumin and a protein-denaturing agent and/or a compound
CC having S-S bond and bromocresol purple for evaluation of glycation
CC proportion of albumin into glycoalbumin with respect to the other
CC albumins measured separately. The compositions are useful in clinical
CC examination including diagnosis and management of diabetes. The
CC compositions can be used in methods which avoid the effects of globulin
CC and ascorbic acid components and stabilise enzymes such as protease
CC acting on glycoamino acids. This sequence represents DNA encoding a
CC polypeptide used in the scope of the invention
XX
SQ Sequence 1320 BP; 347 A; 333 C; 330 G; 310 T; 0 U; 0 Other;

Query Match 23.5%; Score 333.6; DB 6; Length 1320;
Best Local Similarity 55.2%; Pred. No. 2.4e-100;
Matches 720; Conservative 0; Mismatches 569; Indels 15; Gaps 3;

QY 41 TCACTAAGTCATCATCTCTCTGATCGTTGGTCCGGGACTTGGGCGACCTCAACGGCTC 100
DB 11 TCACCAACAGTCCCAAAATCTCATCGTTGGTGGCGAACTTGGGGATGCTCAACTGCC 70
QY 101 TGCACCTTCGGCGCGCGGGATATACCAACGTTACCGTGTGGACCCCTATCTCTGCCCTA 160
DB 71 TCCATCTCGCCCGTGGGGTTACCAACGTCACCTGTTCTCGATGCAATCGCATCCCGT 130
QY 161 GCGCCATCTCCCGCGGAAACGAGTGAACAAAGTCATTAGCAGTGGCCAAATATTCGAATA 220
DB 131 CACCGATATCAGCGGGCATGATGTAAACAACTTGTGCGCCGACTGTGCGACTGCCGATA 190
QY 221 ACAA-----GAGGAATCGAAGTGAATGAGATCTTGGCGGAGAGCGGTTTAAACGGTT 274
DB 191 GCAAGGTGATGATGAAGACTCAATCTGGAAGCACTTAGCTACGCCGCGACGCTCAAGGAT 250
QY 275 GGAAGAACGACCGCTTTTCAACCGTATTATCATGATACGGCTCTGCTGATGCTGCTT 334
DB 251 GGTCTCAGACCTCTGCTTCCACCATCTGCCAATACAGGCTCTGTCGTGGTGGCT 310
QY 335 GCTCGCAGGAGGCGCTGGATCGCTGCGCGCTCGGGTACGTCGCGGCGAGGATCCTAATC 394
DB 311 CAACACCAAGTCTATCAACGAGCTGTGTAAGATGAGATCGGTGACGACATCGACCGT 370
QY 395 TGGTGAACCTTACCGCCCGGAGCAATTCGTAACTGGCCCGGAGAGCGGTGTGCAAG 454
DB 371 ATACACCTCTCAACACAGCAGAGAATTTCAAGAAAGCACCATGCTCGAGGGTATCTCGAC 430
QY 455 GTGATTTTCGGGTGGAAAGGTACTTTTCGGGTTCGGCGCTGCGTGGGCGACATGCAA 514
DB 431 GTAACTTTCAGGCTGGAAGGGCTTTTAAAGCCACGCGGTTCTGGTGGGTTCATGCTC 490
QY 515 GGAATGCTTTAGTGGCAGCAGCAGCGAAGCAGCGCATGCGGTGTAAATTTTGTACTG 574
DB 491 GAAAGCTATGAAGCTGCTTTTCGAGAGAGCGAGAGCTTGGTGTCAATTCATCACTG 550
QY 575 GCAACCGCGAGGTGCTGTAGTCAAGTAACTTTTGAATAATACGATGTAAGAGGTGCCG 634
DB 551 GCTCTCCGAAGGAAAGGTGGAGAGTCTGATCTTTGAAGACGCGCATGTTTCGAGGTGCCA 610
QY 635 TTACGGCGGATGCAAAATTTGGAGAGCGCAACCTATCTCTGCTGTGGCGGTAGCG 694
DB 611 AGACGGCAGATGTTAAGGAGCAGCAGAGCGGATGAACTATCTTCGCTGGGTCTCAG 670
QY 695 CGGTCAGTTTCTAGATTTTCAAGAAATCAACTTGCACCAACCGTGTGACCCCTGGTACACA 754
DB 671 CAGAGTTCTTCTCGATTTTGGAGAACAGATCCAGCTACGGCGTGAACCTTGGCCATA 730
QY 755 TTGGTTTAAACCGGAGAAAGTGGTGTGTAACAAATATACCGGTTTATCTTTAAACATCG 814
DB 731 TCCAGATGACACCAAGAGAAACCAAGCTGTACAGAACTTGCACACCTCTTTTCAACATCA 790
QY 815 AACGGGGTTTCTTGTGAACCGGATGAGCGCGGTGAGATTAAATATGCGATGAAC 874
DB 791 ACCAAGGTTTCTTATGGAAACCTGATGAGGATCTTTCATCAACTCAAGATGTGGGATGAAC 850

QY 875 ACCCGGGCTACACAAATATGTTCCAGAGTCCAGAGCGCAGC---ATGATGAGCATTCGGT 931
DB 851 ATCCGGGCTACTGCTCAACTGGTGTGAAAGAGCTGGTCTTAAGTACCCCGAGTCCATCCCT 910
QY 932 TCCAAAGAAACCCAGATTTCACAAAGAGAGCCGAAACCGCGCTTCGGGCGCTTCTGTAAGAGA 991
DB 911 TCCAAAGCAATCAAGTGCACCGAGGCTGAACGACGATGAGAGCAGTTTCTGAAGATA 970
QY 992 CAATGCCCGCAGCTGGCAGACCGTCCATTTCAGCTTCGACGCAATTTGCTGGTGTCCGATA 1051
DB 971 TCATGCTCAGCTTGCAGATCGGCGCTTGTTCATGCTCGAATCTGCTGGTGGCTGATA 1030
QY 1052 CCGCGAATCCGAATTCCTGATAGATCGACATCCGAGTACCACAGCTTGTGTGGCT 1111
DB 1031 CACAGGATAGAAATGTTCTCTGATCACCCTATCATCTCGACATCCCTCACTTGTCAATTGCTT 1090
QY 1112 GTGCTGCGAGCGAAGAGGCTTTAAATATCTGCTTCTTATTTGGAAATCTCATTTGACG 1171
DB 1091 CAGTGATTCGGCAGCGGTTATCATGATATCATCAATTTGGAAGTTTATCTCTGACT 1150
QY 1172 CGATGAAGGTAAAGTCCCGCAAAATTCACGAATTAATCAAGTGGAAACCCGACATTTG 1231
DB 1151 GTATGAGGCTACGCTTGAGGAAGGTTTCCCAAGTTCTCGAGATGCGCAGAGAGT 1210
QY 1232 CGCGCAACCGTAATCTGCGGTGATCTCTGCGGCGTTCGCGGTCGCTGCTGATGATGG 1291
DB 1211 TTACCGAGTTCTGGGTTAAAGATCTCTCGATCGGTTGGAGCTGACGATAAGATCATGG 1270
QY 1292 ATTTTC-----ATGATGTCGAAGGATGGACCAATGTCAGTAT 1329
DB 1271 ATTTGCCAAGAGTGTAGAGGGATGGGAAATATCAAGAAT 1314

RESULT 7

ABK90476
ID ABK90476 standard; DNA; 1320 BP.

XX ABK90476;

XX 05-NOV-2002 (first entry)

DE Fusarium oxysporum polynucleotide #2.

XX Glycoamino acid; glycoprotein; deoxycholic acid; deoxycholic amide; gene;
KW cholic acid amide octyl glucoside; quaternary ammonium salt; ds;
KW quaternary ammonium salt cationic surfactant; concanavalin A; betaine;
KW ascorbic oxidase; 4-(2-hydroxyethyl)-1-piperazinyl group; albumin;
KW protease; glycoalbumin; bromocresol purple; glycation; diabetes;
KW ascorbic acid; mutant.

XX Fusarium oxysporum.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1. .1320
FT /*tag= a
FT /product= "Fusarium oxysporum polypeptide #2"
FT /partial

FT /notes= "No start or stop codon given"
FT /transl_except= (pos:736. .736, aa:lle)
FT /transl_except= (pos:1114. .1116, aa:Trp)
FT /transl_except= (pos:1186. .1188, aa:Tyr)

FT mutation

FT mutation
FT replace(1115,A)
FT /*tag= b
FT /tag= c

XX WO200261119-A1.

XX 08-AUG-2002.

XX 30-JAN-2002; 2002WO-JP000721.

PF

XX	31-JAN-2001; 2001JP-00022953.	
PR	16-FEB-2001; 2001JP-00039796.	
PR	08-AUG-2001; 2001JP-00240002.	
XX		
PA	(ASAH) ASahi KASEI KOGYO KK.	
XX		
PI	Kouzuma T, Yoshioka I, Arai M, Sumitani J, Imamura S;	
XX		
DR	WPI; 2002-599854/64.	
DR	P-PSDB; ABG30783.	
XX		
PT	Compositions for assaying glycoprotein without interference from globulin	
PT	and ascorbic acid in blood, useful in clinical examination including	
PT	diagnosis and management of diabetes.	
XX		
PS	Example 21; Page 66-69; 82pp; Japanese.	
XX		
CC	The invention relates to compositions for assaying glycoprotein with use	
CC	of protease and enzymes acting on glycoamino acids in the presence of at	
CC	least 1 of deoxycholic acid or amide, cholic acid amide octyl glucoside,	
CC	quaternary ammonium salt, quaternary ammonium salt cationic surfactant,	
CC	concanavalin A and betaine and/or ascorbic oxidase and a buffer without a	
CC	4-(2-hydroxyethyl)-1-pyrazolyl group. The invention also relates to	
CC	compositions for assaying albumin containing the protease and enzymes for	
CC	determining glycoalbumin and a protein-denaturing agent and/or a compound	
CC	having S-S bond and bromocresol purple for evaluation of glycation	
CC	proportion of albumin into glycoalbumin with respect to the other	
CC	albumins measured separately. The compositions are useful in clinical	
CC	examination including diagnosis and management of diabetes. The	
CC	compositions can be used in methods which avoid the effects of globulin	
CC	and ascorbic acid components and stabilise enzymes such as protease	
CC	acting on glycoamino acids. This sequence represents DNA encoding a	
CC	polypeptide used in the scope of the invention	
XX		
SQ	Sequence 1320 BP; 346 A; 333 C; 330 G; 311 T; 0 U; 0 Other;	
	Query Match 23.4%; Score 332; DB 6; Length 1320;	
	Best Local Similarity 55.1%; Pred. No. 8.4e-100;	
	Matches 719; Conservative 0; Mismatches 570; Indels 15; Gaps 3;	
QY	41 TCACATAAGTCATCATCTCTCTGATCGTTGGTGC CGGACCTTGGGGCACCTCAACGGGCTC 100	
DB	11 TCACCAACACAGTCCCAAATCTCATCGTTGGTGGCGGAACCTTGGGGATGCTCAACTGCCC 70	
QY	101 TGCACCTCGCGCGCGCGGATATACCAACGTTACCGTGCTGGACCCCTATCTCTCCCTA 160	
DB	71 TCCATCTCGCGCGCGCGGGTTACACCAACGFCACCTGTTCTCGATGTCAATCGCATCCCGT 130	
QY	161 GCGGCATCTCGCGCGGAAACGACGTGAACAAAGTCATTAGCAGTGGCCAAATATTCGAATA 220	
DB	131 CACCGATATCAGCCCGGCATGATGTAAACAAACTTGTGGCCGACTGTCGATCCGATA 190	
QY	221 ACRAA-----GACGAATCGAAGTGAATGAGATCTTTGGCGGAAGAGCGGTTTAAACGGTT 274	
DB	191 GCRAAGGTGATGATGAAGACTCAATCTGGAAAGCACCTAGCTACGCCGACGCTCAAGGAT 250	
QY	275 GGAAGAAGCAGCCGCTTTTCAACCGGTATATCATGATACGCGGCGCTGCTGATGTCGTGTT 334	
DB	251 GGCTCCACGACCCGTGTCTTCCAACCAATCTGCCCAATACACAGGCTCTGTCTCGTGGCTGCT 310	
QY	335 GCTCGCAGGAGGGCGCTGGATCGCTCGTGGCGTCCGGTACGTCCGGCGGAGGATCCTAATC 394	
DB	311 CAACACCAAAAGTCTATCAAGCAGCTGGTGAAGATGAGATCGGTGACACATCGACCCAGT 370	
QY	395 TGGTGGAACTTACCCGCCGCGAGCAATTTTCGTAAACTGGCCCCCGGAAGCGCTGTTGCAAG 454	
DB	371 ATACACCTCTCAACACAGCAGAAGATTCAGAAAGACCATCGCTGAGGGTATCTGCACAG 430	
QY	455 GTGATTTTCGGGGTTGGAAGGGGTACTTTTGGCGGTTCCGGCGCTGGGTGGGCACATGCAA 514	
DB	431 GTAACCTTTTCAGGCTGGAAGGGCTTTTACAAGCCCAACGGGTCTCTGGTTGGGTTATGTC 490	

QY 1292 ATTTTC-----ATGATGTGAAGGAATGGACCAATGTTCAAGTAT 1329
|||||
Db 1271 ATTTGCCAAGAGTGATGTAGAGGGATGGACAAATATCAAGAT 1314

RESULT 9

ADM78788

ID ADM78788 standard; DNA; 1335 BP.

XX

AC ADM78788;

XX

DT 01-JUL-2004 (first entry)

XX

Fusarium proliferatum fructosylamine oxidase Q2 coding sequence.

DE

db; gene; fructosylamine oxidase; fructosyllysine; fructosylvaline;
Amadori compound; medical examination; diagnosis; serum glucose level;
food inspection.

KW

KW

XX

OS Fusarium proliferatum.

XX

PH Key

FT CDS

FT Location/Qualifiers

FT 1.1335

FT /*tag= a

FT /product= "fructosylamine oxidase Q2"

XX

PN WO2004029251-A1.

XX

PD 08-APR-2004.

XX

PF 16-SEP-2003; 2003WO-JP011766.

XX

PR 24-SEP-2002; 2002JP-00277214.

XX

PR 24-OCT-2002; 2002JP-00309734.

XX

XX (ARKR-) ARKRAY INC.

XX

XX Yoshida N, Tani Y, Yonehara S;

XX

XX WPI; 2004-316127/29.

DR

DR P-PSDB; ADM78789.

XX

XX Fusarium proliferatum-originated fructosylamine oxidase useful in

PT

PT measurement of Amadori compound and particularly applicable in medical

PT

PT examination including diagnosis and management of diabetes.

XX

PS Claim 8; SEQ ID NO 5; 50pp; Japanese.

XX

XX The invention relates to a novel fructosylamine oxidase isolated from
XX Fusarium proliferatum. The enzyme has the following physicochemical
XX characteristics: (a) activity against fructosyllysine equivalent to or
XX higher than that against fructosylvaline; (b) optimum pH for enzyme
XX activity of 7.5; (c) enzyme stability optimally at 30-40 deg C; and (d) a
XX molecular weight of about 39 kDa as measured by SDS-PAGE and about 39.4
XX kDa by gel filtration. Also disclosed is a similar Fructosylamine oxidase
XX characterized by: (i) activity against fructosylvaline and no detectable
XX activity against fructosyllysine; (ii) optimum pH for enzyme activity of
XX 7; (iii) enzyme stability optimally at 30-40 deg C; and (iv) a molecular
XX weight of about 49 kDa as measured by SDS-PAGE and about 58 kDa by gel
XX filtration. The oxidases are useful in the measurement of Amadori
XX compounds and particularly applicable in medical examination including
XX diagnosis and management of diabetes by controlling serum glucose level
XX for treatment and prevention of complications, and in food inspection.
XX This sequence represents the coding sequence for the fructosylamine
XX oxidase designated Q2.

SQ

Sequence 1335 BP; 363 A; 328 C; 333 G; 311 T; 0 U; 0 Other;

XX

Query Match

Best Local Similarity 22.9%; Score 324.4; DB 12; Length 1335;

Matches 696; Conservative

0; Mismatches 571; Indels

9; Gaps

2;

QY 31 GCTCTCGCGTCACTAAGTCATCATCTCTCTCGATCGTTGGTGCCGGGACTTTGGGGCACC 90
Db 13 GTTGGCCCGCTCAATAAGGACTCAGGGAATCTCATCGTTGGTGGCGAACTTGGGGATGC 72
QY 91 TCNACGGCTCTGCACCTCGCGCGCGCGGATATACCAAGTTACCTGCTGGAGCCCTTAT 150
Db 73 TCNACTGCCCTCCATCTCGCCCGTCTCGGGGTACACCAACGTCACTGTTCTCGATGTCAT 132
QY 151 CTTGTCCTTAGCGCCATCTCCGCCGGAACGAGCTGCAACAAAGTCATTAGCAGTGGCCAA 210
Db 133 CGATCCCGTACCAGATATCAGCCGGGCGATGATGTAACAACAATTTCTTAACACTAGGC 192
QY 211 TATTCGAATAACAAGACGAATTCGAAG-----TGAATGAGATCTTGGCGGAAGAGCG 264
Db 193 ACTTCTGATAGTAAAGCGGATGACGAAGACTCAATCTGGAAGCTCTTACGTACGCCGCA 252
QY 265 TTTAACGGTTGMAAGAACGACCGCTTTTCAAAACGTTATATCATGATACGGGCTGCTG 324
Db 253 GCTCAAGGATGGCTCCATGATCCCATCTTCCAACCTTTTCGCCACAATACAGGAGCTGC 312
QY 325 ATGTCCTGCTTCTCGCAGGAGGCGCTGGATCGCTCGGGGTCCGGGTACGTCCTCGGGCGAG 384
Db 313 ATGGCTGGCTCAACACAAATCTATCAGCAGCTGTTAGAGATGAGATCGGTGAGCAC 372
QY 385 GATCTTAATCTGTTGGAATCTTACCCCGCCGGAGCAATTTTGTAAACTGGCCCCGGAAGGC 444
Db 373 ATCGACCACTATACACCTCTCAACACAGCAGAAGATTTTCAAGAGGACTATGCCGGAGCT 432
QY 445 GTGTTCAAGGTGATTTTTCGGGTTCGAAAAGGTACTTTTTCGGGCTTCCGGCGCTGGCTGG 504
Db 433 ATTCTGACAGGTGATTTTCTAGGCTGGNAGGGCTTTTACAAGCCCAAGAGGTTTCAGGTGG 492
QY 505 GCACATGCAAGGAATGCTTTAGTGGCAGCAGCACGCGAAGACACAGCGCATGGGTGTAAAA 564
Db 493 GTTCAITGCCAAGAAAGCTATGAAGCTGTTTGAAGAGAGCCAGAGACTTGGTGTCAAG 552
QY 565 TTTGTTACTGGCACCCCGCAGGCTCGTGTAGTCACGTTAATCTTTGAAAATAACGATGTA 624
Db 553 TTTCACTACCTGGCTCTCCCGAAGCAAGGTCGAGAGTCTGGTCTTTGAAAGCTGGTGTGTC 612
QY 625 AAGGTGCGGTACGGGCGATGSCAAAATTTGGAGAGCGGAACGTACATCTCTGTGTGCT 684
Db 613 AAGGTGCAAAAACACAGCATGGAAGAAACACAGAGCGGATCGAACAAATTTCTCTCGCT 672
QY 685 GGGGCTAGCGCGGCTCAGTCTCTAGATTTCAAGAACTCACTTCGACCAACCGCTTGGACC 744
Db 673 GGTGCTCAGCAGAGTCTCTCCCTCGATTTTGAACACAGATCCGCTCTACGGATGACT 732
QY 745 CTGGTACACATTCGCTTAAACCCGGAAGAACGTCGCTTGTACAAAATAATATACCGGTTATC 804
Db 733 CTGGGCGCATATCCAGATGACAGCAGAGGAAACAAAGCTGTACAAAGAACTTTCCCCCGCTT 792
QY 805 TTTAATCATGAACGGGGTTTTTCTTTTGAACCGGATGAGGAGCGCGGTGAGATTAATAA 864
Db 793 TTTCAATATCAACACAGGGCTTCTTCATGGAACCCGATGAGGACTTGCATCAACTCAAGATG 852
QY 865 TGGATGAACACCCCGGCTTACACAAATATGTTCCAGAGTGCAGACGCA---CGATGATG 921
Db 853 TGGATGAACATCCCGGATCTGCAATTTGGTTGACAAACCTGGTTCCAAATACCCCGAG 912
QY 922 AGCATTTCCGTTGAAAACCCAGATTTCCAAAAGAACCGGAAACGCGGCTTCGGGGCGCTG 981
Db 913 TCCATCCCTTCGAAAAGTATCAAGTGCCTCAATTTGAGGCTGAAACGAGCATGAAGCAATTT 972
QY 982 CTGAAGAGACATGCCCCAGCTGGCAGACCGTCCATTAGCTTCGACCGCATTTGCTGG 1041
Db 973 CTGAAGAGACATCATGCTCAGCTCGCAGATCGGCCCACTTGTTCATGCTCGAATCTGCTGG 1032
QY 1042 TGTGCGGATACCGCGAATCGCAATTTCTGTAGATCGACATCGGAGTAGTACCAAGTCTT 1101
Db 1033 TGGCCGATACACAGATAGATGTTTCTGATCAGTATCACCTCGACACCCATCGCTT 1092
QY 1102 GTGTTGGGCTGTGGTCGAGCGGAAGGGGTTTAAATATCTGCTTCTATTGGAATCTC 1161

Db 1093 GTCAATGCTTCCGGGATTTGGGCACAGGATACAGCATATCACTTTCATTGGAAAGTTC 1152
 Qy 1162 ATTGTTGACGCGATGGAAGTAAAGTCCGCCAAAAATTCACGAATTAATCAAGTGGAAAC 1221
 Db 1153 ATCTCTGATTGATGAGGGGCACATTGGAGGAAAGTTTGTAAAGTTCTGGAGATGGCGA 1212
 Qy 1222 CCGGACATTCGGGCGAACCGTAACTGCGGTGATACCTGGGGCGTCTTTGGGGTCCAAAT 1281
 Db 1213 CCAGAGAAAGTTTACGGAGTTCTGGGGTAAAGATCCCTGGATCGGTTTGGAGCTGACGAT 1272
 Qy 1282 CGTGTGATGATTTTC 1297
 Db 1273 AAGATCATGATTTGC 1288

RESULT 10

ADS15772

ID ADS15772 standard; DNA; 1284 BP.

AC ADS15772;

16-DEC-2004 (first entry)

Fructosylamine oxidase coding sequence, SEQ ID 2.

Fructosylamine oxidase; enzyme; fructosylamine; fructosyl valine; gene;

ds.

Pichia sp.

Key Location/Qualifiers

CDS 1..1284

/*tag= a

/product= "Fructosylamine oxidase"

WO2004083419-A1.

30-SEP-2004.

17-MAR-2004; 2004WO-JP003587.

17-MAR-2003; 2003JP-00116348.

(SODE/) SODE K.

Sode K;

WPI; 2004-709934/69.

P-PSDB; ADS15771.

Novel fructosylamine oxidase protein, useful for spectroscopic analysis of fructosylamine compounds, and electrochemical analysis of fructosylamine and fructosyl valine compounds.

Claim 6; SEQ ID NO 2; 51pp; Japanese.

The present invention relates to fructosylamine oxidase protein (I; ADS15771) and its coding sequence (II; ADS15772) from *Pichia* sp. preferably *Pichia* sp. N1-1 strain. (I) is useful for spectroscopic analysis of fructosylamine compounds, and electrochemical analysis of fructosylamine and fructosyl valine compounds. (I) is also useful for analyzing fructosyl valine by spectroscopy, which involves degrading HbA1c, fructosamine or glycoalbumin in a sample using (I) to produce fructosyl valine or fructosylamine, respectively. (I) is also useful for carrying out electrochemical analysis of HbA1c, fructosamine, and glycoalbumin.

Sequence 1284 BP; 391 A; 253 C; 301 G; 339 T; 0 U; 0 Other;

Query Match 8.18; Score 115.2; DB 13; Length 1284;

Best Local Similarity 47.4%; Pred. No. 5.1e-27;

Matches 524; Conservative 0; Mismatches 558; Indels 24; Gaps 5;

Qy 55 TCTCTCTGATGCTTGGTGGCGGACCTTGGGGCACCTCAACGGCTCTGCACCTCGCGGC 114
 Db 7 TCATATAATTAAGTTGGTGGCGGTACTTTTGGGCTTCCACAGCTTACAGCTTGGCCAGA 66
 Qy 115 CGGGATATACCAAGCTTACCGTGTGGACCCCTATCTCTGCTCCCTAGCGCATCTCGGCC 174
 Db 67 GATGATACAAGAACATAAATGTTTGGACAGTTTCGGTTCATCTGAGTAGCTGCT 126
 Qy 175 GGAACGACGTGAACAAAGTCATTAGCAGTGGGCAATATTCGAATAACAAGACGAATC 234
 Db 127 GGAACGACAGTAAACAAGATTTTTCAC- - -TACGATTATGTGCTCCCTGGCTAAACCC 183
 Qy 235 GAAGTGAATCAGATCTTGGCGGAGAGCGCTTAAACGGTTGGAAGACGACCCGCTTTC 294
 Db 184 AATTCAAAAGAACGGTTGAGTCTCGAAGCATTAACCTTTTGGAAAGACAGATCCGGTGTAC 243
 Qy 295 AAACCGTATTATCATGATACGGGCTGCTGATGCTGCTTGTCTGCTCGCAGAGGG- - -CCTG 351
 Db 244 AAACCGTACTATCATCGGTAGGATTTATCTGCTGCAAGTTCGGATGCTCCATTACTG 303
 Qy 352 GATCGCTGGCGTCCGGGTACGTCGGGGGAGATCCTAATCTGGTGGAACTTACCCGC 411
 Db 304 CATGATAAGGAATACTATGAAGAGTTGCAAAAACGACTTCGCAATTAATCGTTATATT 363
 Qy 412 CCGAGCAATTTCTGTAACCTGGCCCGGAAAGGCGTGTGCAAGTGAATTTTCCGGTTGG 471
 Db 364 TCAACTCCCGAGGAGTTTTCGTGAGTATTTGCGCATTTTAAAGGGCCGCTTACCCAACCTG 423
 Qy 472 AAAGGTACTTTTCGGGCTTCGGCGCTGCGGCATCGGCACATCGAAGAAATGCTTAGTGCA 531
 Db 424 AGAGGATATGTTCTCGACGGAGATACGAGTGTGTCATGCTCGAGACTCATTTGAAAGT 483
 Qy 532 GCACGACGCGAAGCACAGCGCATGGGTGTAATAATTTGTTACTTGGCACCCCGCAGGGTGT 591
 Db 484 GCATACGAAGATGCAAAACGATTGGGAGTGAATTTGTTTGGAGAC- - -GATGGGAA 540
 Qy 592 GTAGTACAGTTAATCTTTGAAATACGATGTAAGTGTGCGGTGCGGTAGCGGCGATGGCAA 651
 Db 541 ATTGTGGAATTAATTAACGAAATGGAAGTTGACGGGAAATTAGGGCCAGATCTGGTGCC 600
 Qy 652 ATTTGGAGAGCGGAACGTACATCTCTGTGTGCTGGGCTAGCGGGTCAAGTTCCTAGAT 711
 Db 601 ATATTTCCGACAAAAAATATGTTCTCAGCTCTGGTGCAAAATGCAAGTGAACGTTGTTAAT 660
 Qy 712 TTCAAGAAATCAACTTCGACCAACCGCTTGGACCCCTGGTACACATTTGGGTTAAAAACCGAA 771
 Db 661 TTCCAGAGACAGCTAGAAGGTAAATGTTTCACTTTGGCACATTTCAAAGTGACGATGAA 720
 Qy 772 GAACGTGCTGTGTACAAAAATATACCGGTTATCTTTAAACATCGAACCGGGGTTTCTTTT 831
 Db 721 GAAGCTAAAGCAATTTAAAGCTTGGCGTCTCTTTTCAATGCCGAAAAAGGGTTTTTTTC 780
 Qy 832 GAACCCGATGAGGCGCGGTGAGATTTAAATATGCGATGAACACCCGGGCTACACAAAT 891
 Db 781 GAGGCTGATGAAATAAC- - -GAAATCAAAATTTGCAACGAGTACCCCTGGATTACCCAC 837
 Qy 892 ATGGTCCAGAGTGCAGACGCGACGATGATGAGCATTCGGTTCCGAAAAAACCCAGATTCCA 951
 Db 838 ACAAATGAATCCGGAGA- - -GTCTATCCACTCTACCGGATGGAGATTCCA 885
 Qy 952 AAAGAAGCGAAACGCGGCTTGGGGCCCTGCTGAAAGAGACAAATGCCCGAGCTGGCAGAC 1011
 Db 886 CTCGAGTCAGCACTTGAATTAGACAATACTTGAAGAAACCAATGCTCAGTTTGTGCTGAT 945
 Qy 1012 CGTCCATTGAGCTCGACGCAATTTGCTGGTGTGCGGATACCGGAATCGGGAATTCCTG 1071
 Db 946 AGACCTTTTCAACAAGACAAAGAAATTTGTGTGTACCGACTCTCCCGACATTCGAATTGATC 1005
 Qy 1072 ATAGATCGACATCCGAGTACCAAGTCTTGTGTGGGCTGTGTGGCGAGCGAAGAGGG 1131
 Db 1006 TTGTGTACTCCACGAGATACCAACCTTATTTGTAGATCGGCTGACTCTGGAAATTCG 1065


```
QY 1184 AAGTCCGCAAAAATTCAGAAATTAATCAAGTGGAAACCCGGACATTTGCGGGAACCGTA 1243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1187 CTCATCGCAGGAATGGCTGGTCTCGGATGGAGATGGAGACCCGGAGGTGATGCTCTTAGAT 1246

QY 1244 ACTGGCGTGATCTCTGG 1261
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1247 CTAGACGCGGTGCTCCGG 1264

RESULT 14
AA15949
ID AAX15949 standard; DNA; 1314 BP.
XX
AC AAX15949;
XX
DT 14-MAY-1999 (first entry)
XX
KW DNA encoding a fructosylamino acid oxidase.
DE
XX Fructosylamino acid oxidase; FAOD-P; Penicillium janthinellum S-3413; ds.
XX
OS Synthetic.
OS Penicillium janthinellum.
XX
FH Key Location/Qualifiers
FT 1..1314
FT CDS /*tag= a
XX
PN JP11046769-A.
XX
PD 23-FEB-1999.
XX
PF 05-AUG-1997; 97JP-00210609.
XX
PR 05-AUG-1997; 97JP-00210609.
XX
PA (KYOT-) KYOTO DAIICHI KAGAKU KK.
XX
WPI; 1999-208112/18.
DR P-PSDB; AAW97386.
DR
PT New DNA coding fructosylamino acid oxidase - synthetically designed and
PT recombinantly prepared.
XX
PS Claim 1; Page 8-10; 21pp; Japanese.
XX
CC The present DNA sequence encodes a fructosylamino acid oxidase (FAOD-P).
CC The present sequence is a synthetic sequence, derived from the wild type
CC FAOD-P sequence of Penicillium janthinellum S-3413. The sequence was
CC constructed using primers AAX15951-6006
XX
SQ Sequence 1314 BP; 408 A; 168 C; 244 G; 494 T; 0 U; 0 Other;

Query Match 4.1%; Score 57.6; DB 2; Length 1314;
Best Local Similarity 54.3%; Pred. No. 1.1e-07;
Matches 139; Conservative 0; Mismatches 114; Indels 3; Gaps 1;

QY 628 GGTGCGGTACGGCGATGCGCAAAATTTGGAGAGCGGAACGTACATTCCTGTGCTGGG 687
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 604 GGTGTTGAAACTGTTGATGCTGCTACTAAATATTTTGCTGATAAAGTTGTTTAGCTGCTGGT 663

QY 688 GCTAGCGCGGTGAGTTCTTAGATTTCAAGAACTCAAGTTCGACCAACCGCTTGGACCGTG 747
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 664 GCTTGGTCTTCTACTTTAGTTGATTTAGAGATCAATGTGTTTCTAAGGCTTGGGTTTTT 723

QY 748 GTACACATTCGGTTAAACCGGAAGACGGTGGTTGTACAAAATATACCGGTTATCTTT 807
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 724 GCTCATATACAATTAACCTCCACAAGAACTGCTCAATATATAAAGATGTTCCAGTTGTTAT 783

QY 808 AACATCGAAGCGGGTTTTTCTTTGACCCGATGAGCGCGGTGAGATTAAATATGC 867
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 784 GATGGTGATTATGGTTTTTTTTTTTGAACCAAAAT---GAACATGGTGTATATAAAGTTGT 840
```

```
QY 868 GATGAACACCCGGGCT 883
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 GATGAATTTCCAGGTT 856

RESULT 15
AAT85703
ID AAT85703 standard; cDNA to mRNA; 1314 BP.
XX
AC AAT85703;
XX
DT 16-FEB-1998 (first entry)
XX
DE Coding sequence for fructosyl amino acid oxidase.
XX
KW Fructosyl amino acid oxidase; FAOD-P; enzyme; amadori compound;
KW alpha-ketoaldehyde amine derivative; amadori detection; ss.
XX
OS Penicillium janthinellum.
XX
FH Key Location/Qualifiers
FT 1..1314
FT CDS /*tag= a
XX
PN WO9721818-A1.
XX
PD 19-JUN-1997.
XX
PF 13-DEC-1996; 96WO-JP003651.
XX
PR 14-DEC-1995; 95JP-00325672.
XX
PA (KYOT-) KYOTO DAIICHI KAGAKU CO LTD.
XX
PI Kato N, Sakai Y, Tani Y, Fukuya H;
XX
WPI; 1997-332789/30.
DR P-PSDB; AAW24134.
DR
XX Fructosyl amino acid oxidase which oxidises amadori compounds and related
PT DNA - used for detection of amadori compounds in foods and in body fluids
PT such as blood.
XX
PS Claim 1; Page 33-36; 53pp; Japanese.
XX
CC This sequence represents the coding sequence for the fructosyl amino acid
CC oxidase (FAOD-P) of the invention. This sequence was isolated from
CC Penicillium janthinellum S-3413 (FERM BP-5475). FAOD-P oxidises amadori
CC compounds in the presence of oxygen, producing alpha-ketoaldehyde, amine
CC derivatives and hydrogen peroxide. FAOD-P is used in the detection of
CC amadori compounds in, e.g. foods such as soy sauce, and in body fluids
CC such as blood. This sequence, vectors containing it and host cells
CC transformed by the vector are all used for the recombinant production of
CC FAOD-P
XX
SQ Sequence 1314 BP; 339 A; 323 C; 345 G; 307 T; 0 U; 0 Other;

Query Match 3.8%; Score 53.6; DB 2; Length 1314;
Best Local Similarity 53.1%; Pred. No. 2.5e-06;
Matches 138; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

QY 628 GGTGCGGTACGGCGATGCGCAAAATTTGGAGAGCGGAACGTACATTCCTGTGCTGGG 687
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 604 GGTGTTGAGACAGTACATGGAACAAATACTTCGCCGACAAGGTGTTTGGCGCGCTGTT 663

QY 688 GCTAGCGCGGTGAGTTCTTAGATTTCAAGAACTCAAGTTCGACCAACCGCTTGGACCGTG 747
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 664 GCTTGGAGTTCCAGCTGTAGTAGATTGGAGGACCAATGTGTTTCGAAGCGCTGGGCTTC 723

QY 748 GTACACATTCGGTTAAACCGGAAGACGGTGGTTGTACAAAATATACCGGTTATCTTT 807
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 724 GCTCATATCCAACCTACGCCCCCAAGAAATCGGCCAGTACAGGACGTGCCCGTAGTATAC 783
```

Qy	808	AACATCGAACGGGGTTTTTTTCTTTTGAACCCGATGAGGAGCGCGTGAGATTAAATATGC	867
Db	784	GACGGTATTATGCTTTTCTTCGAGCCC---AACGACACGGAGTAATCAAGTCTGC	840
Qy	868	GATGAACACCCCGGCTACAC	887
Db	841	GATGAGTTCCCCCGGTTCTC	860

Search completed: May 29, 2005, 20:15:46
Job time : 844 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2005, 19:21:38 ; Search time 5114 Seconds
(without alignments)
10561.830 Million cell updates/sec

Title: US-10-622-893A-6
Perfect score: 1419
Sequence: 1 atggggggttcgggtgacga.....atccaccatcatcatcattaa 1419

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	237.4	16.7	782	7	CO142686 EST837357
C 2	118.4	8.3	1040	9	CNS07A13
3	99.2	7.0	906	9	CNS07AGE
4	83.4	5.9	440	1	AA787493
C 5	72.6	5.1	397	4	BI187569
6	66.2	4.7	677	7	CN241854
7	65.2	4.6	414	4	BI187570
8	64.6	4.6	599	1	AJ636269
9	64.6	4.6	628	4	BM869701
C 10	63.4	4.5	459	7	CF188389
11	57	4.0	422	4	EG465351
12	55.2	3.9	758	7	CF871025
13	55.2	3.9	816	6	CB901239
14	55.2	3.9	822	6	CB901049
15	55.2	3.9	822	7	CF870843
16	55	3.9	613	7	CK907858
17	53.8	3.8	491	7	CF188390
C 18	50.2	3.5	878	9	CNS07990
C 19	50	3.5	795	7	CF684496
C 20	49.6	3.5	498	1	AA901536
21	45.8	3.2	686	8	AQ447867
22	44.6	3.1	375	1	AJ237240
23	43.4	3.1	686	7	CF883318
C 24	43.4	3.1	756	6	CB909424

25	43	3.0	486	7	CK909112
26	42.6	3.0	746	7	CF684508
C 27	42.6	3.0	749	7	CF718110
28	40.8	2.9	794	9	CC997434
C 29	40.6	2.9	1009	9	CC782545
30	40.4	2.8	934	9	CNS07BFZ
31	40	2.8	551	7	CO056263
32	40	2.8	598	7	CO053986
C 33	39	2.7	683	6	CD309225
34	38.4	2.7	438	6	CA962243
35	38.4	2.7	468	7	CK731051
36	38.2	2.7	349	5	BQ497928
C 37	38.2	2.7	1255	8	CC309899
38	38	2.7	337	2	AW309033
39	38	2.7	431	6	CB025818
40	37.6	2.6	606	7	CF573978
41	37.6	2.6	833	9	CC838405
42	37.4	2.6	1101	9	CNS00370
C 43	37.2	2.6	436	6	CD734891
C 44	37.2	2.6	525	1	AA029863
C 45	37.2	2.6	584	5	BX275202

ALIGNMENTS

RESULT 1
CO142686
LOCUS
DEFINITION
EST837357 Aspergillus flavus Normalized cDNA Expression Library
Aspergillus flavus cDNA clone NAFES50 5' end similar to (O42629)
Fructosyl amine:oxygen oxidoreductase, mRNA sequence.
CO142686
CO142686.1 GI:48896687
EST.
Aspergillus flavus
Aspergillus flavus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 782)
Yu, J., Whitelaw, C.A., Nierman, W.C., Bhatnagar, D. and Cleveland, T.E.
Aspergillus flavus expressed sequence tags for identification of
genes with putative roles in aflatoxin contamination of crops
FEMS Microbiol. Lett. (2004) In press
Contact: Yu J
Food and Feed Safety Research Unit
USDA/ARS, Southern Regional Research Center
1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA
Tel: 504 286 4405
Fax: 504 286 4419
Email: juiyu@srcc.ars.usda.gov
Contact Dr. Yu at USDA/ARS SRRC (juiyu@srcc.ars.usda.gov) for clone
information
PCR Primers
FORWARD: M13P
BACKWARD: M13R
Seq primer: M13 Forward
POLYA=No.

FEATURES
source
1..782
/organism="Aspergillus flavus"
/mol_type="mRNA"
/strain="NRRL 3357"
/db_xref="taxon:5059"
/clone="NAFES50"
/sex="asexual mycelia"
/cell_type="mycelia"
/dev_stage="developmental stages from 18 to 96 hours"
/lab_host="E. coli DH10B T1 resistant cells"
/clone_lib="Aspergillus flavus Normalized cDNA Expression Library"
/note="Vector: pBlueScript (SK+) (Stratagene), antibiotic selection marker: Carbenicillin; Site_1: NotI, at the 5

prime end; Site 2: EcoRI, at the 3 prime end; This normalized cDNA expression library was constructed using a mixture of mycelial cells grown under eight different medium conditions and harvested at 5 time points (18, 24, 48, 72, 96 hours). The poly-A sequence was trimmed off before ligating to vector."

ORIGIN

```
Query Match      16.7%; Score 237.4; DB 7; Length 782;
Best Local Similarity 58.8%; Pred. No. 8.8e-63;
Matches 431; Conservative 0; Mismatches 296; Indels 6; Gaps 1;

QY 454 GGTGATTTTCGGGGTGGAAAGGTACTTTGCGCGTTCCGGCGCTCGCTGGCGCACATGCA 513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14 GGCACCTTCCCGGCTGGAAAGGCTGGCTGAACAAGACCGCGCCGGATGGATCCAGCC 73

QY 514 AGGAATGCTTGTAGTCAGCAGCACGACGAAAGCAGCAGCGATGGGTGTAAATTTGTACT 573
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 74 AAGAAGGCCATGTTCTCGGCATACACGGAAGCAAGCGCCTAGGAGTCACTTTTCATCAC 133

QY 574 GGCACCCCGCAGGTCGTGTAGTCAGTTAATCTTTTGAAATTAACGATGTAAAGGTGCC 633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 GGCTCCCTCAAGGACGCTGTATCTCTAATTTACGAGATGGAGACGTAGTCGAGGCC 193

QY 634 GTTACGGGGGATGGCAAAATTTGGAGAGCGGAACGTACATTCCTGTGTCTGGGGCTAGC 693
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 AGAACGGCGAGCGCACCGCTCCACCGCAGCAGCATACCATCTTTCCGCGAGGGCTGGC 253

QY 694 GCGGGTCAGTCTCTAGATTTCAAGATCAACTTCGACCAACCGCTTGGACCTGGTATCAC 753
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 AGTGATCGTCTCTGGACTTTAAGAAACAGCTCCGCTCCTACCGCTGGACGCTCTGCCAC 313

QY 754 ATTGCGTTAAACCGGAGAACGTGGTGTACAAAATATACCGGTATCTTTTACATC 813
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 314 ATCAGAAATGACGCCCGCAGCGAGGCCAAGATCCGGAATCTTCTGTGTTCACAGTC 373

QY 814 GAACGGGGGTTTTTCTTTGAACCCGATGAGGAGCGGGTGAGATTAATAATATCCGATGAA 873
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 374 GCTAAGGGGTTCTTCATGGAACCTGATGAGGATAATCATGAGCTTAAGATCTGCGACGAG 433

QY 874 CACCCGGGTACACAAATATGTTCCAGAGTGCAGA-----CGGCAGATGATGAGCATT 927
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 CATCTCTGATATTGCAACTTCGTCCCGACCGACCGAGCAGCGCGTGAGGTGCGCAGTATC 493

QY 928 CCGTTTCGAAAAACCCAGATTCCAAAGAGCGGAACCGCGCTTCGGGCCCTCTCTGAAA 987
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 494 CCATTTGCAAAAGCATCAGATTCTCTTGAAGCGAGCGCCGTCGAAGGGACTTCCTCCGT 553

QY 988 GAGACAATGCCCGAGCTGGCAGACCGTCCATTCACTTCGACGCGCATTTGCTGCTGTGCC 1047
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 554 GATACGATGCTCATCTTGTGTATCGACCACTGTCTTTTGTCTGTATATGCTGGGATGCT 613

QY 1048 GATACCGCGAATCCGGAATTTCTGTATAGATCGATCGACATCCGCGAGTACCACAGTCTTGTG 1107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 614 GATACAGTGGATCGGGCTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 673

QY 1108 GGTGTGTGTCGAGCGGAGAGGGTTTAATATCTGCTTCTATTCGGATCTCATTTGT 1167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 674 GCTGTGCGTGGATCTGTGTAATGAGCCATGCAAAATGCTTACCATTTGTTGATAGCG 733

QY 1168 GACCGCATGGAAG 1180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 734 GATGCTCTGGAGG 746
```

RESULT 2

CNS07AL3/c
LOCUS
DEFINITION
T7 end of clone XBB0AA002F11 of library XBB0AA from strain CBS 4732
of *Pichia angusta*, genomic survey sequence.
ACCESSION
AL435997
VERSION
AL435997.1 GI:12219410
KEYWORDS
GSS.

```

Pichia angusta
Pichia angusta
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
1 (bases 1 to 1040)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neugeglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
11152876
Pichia angusta
1 (bases 1 to 1040)
Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
and Dujon,B.
Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
angusta
FEBS Lett. 487 (1), 76-81 (2000)
20584723
11152888
Pichia angusta
3 (bases 1 to 1040)
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Creteilux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqlif@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1..1040
/organism="Pichia angusta"
/mol_type="genomic DNA"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone="XBB0AA002F11"
/clone_lib="XBB0AA"
/notes="end : T7"
complement(<331..->972)
/notes="similar to O43029 [ Fructosyl amine; SPBC354.15 ] [
Schizosaccharomyces pombe ]"
/evidence=not_experimental

misc_feature
Query Match      8.3%; Score 118.4; DB 9; Length 1040;
Best Local Similarity 50.5%; Pred. No. 3.1e-25;
Matches 380; Conservative 0; Mismatches 351; Indels 21; Gaps 3;

QY 583 CAGGGTCGTGTAGTCACGTTTAATCTTTTGAAATAATACGATGATAAAGTCCGTTACGGGC 642
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 972 CAGGATACTGCCGAAGAATTGCTCATGGAAGGGGCTCAATTGTCGAGATTGCGACATAT 913

QY 643 GATGGCAAAATTTGGAGAGCGGAACGTACATCTCTGTGCTGGGCTAGCGGGTCAG 702
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 912 TCGGGATCAATCTTGAAGCTGATCGTGTATATATGCTGGAGCAAAATTCACAAA 853

QY 703 TTCTTAGATTTCAAGAATCAACTTCGACCAACCGCTTGGACCTTGATACATTCGCTTA 762
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 852 TTACTCAATTTTAAAGCCCACTACTAGCCAATGCTGGACCGTGGTGCATATAAAGCTT 793

QY 763 AAACCGGAAGAACGTGCTGTGTACAAAATATACCGGTTATCTTTTAACATCGAACGGGG 822
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 792 ACCTCAGAGGCTAAAGCGCTGAAAACTCTCCTGTTCTTTTGAACCTAGTAGAGGG 733
 Qy 823 TTTTCTTTTGAACCCGATAGAGCGCGGTGAGATTAATAATATGCGATGTAACACCCCGGC 882
 Db 732 TTTCTTTTGGAGCCGACGAAAA---TGGGATCCTAAAGTTTGGCAACGAGTTTCCAGGA 676
 Qy 883 TACACAAATATGTTCCAGAGTGACAGCGCGCAGCATGATGAGCATTCGTTTCGAAAAACC 942
 Db 675 TACATAAACCAAGACCGCATGTAATAATTTTCGAGC-----CCCTTTATTCGAAAC 625
 Qy 943 CAGATTTCAAAGAAAGCCGAAACCGCGTTCGCGCCCTGCTGAAAGAGACAAATGCCCCAG 1002
 Db 624 CAGATCCCGTGGAGCGGAAGCAATGAGAGGTTTGTGAAGCAGGTATTTCCCAA 565
 Qy 1003 CTGGCAGACCGTCATTCAGCTTCGACGAGATTTGCTGGTGTGCGATACCGGAATCGC 1062
 Db 564 ATCAGCACCGGTCCTTTGATGCTGCCAAAATTTGCTGGTGCACCGACACACCCGATCGT 505
 Qy 1063 GAATTCCTGATAGATGACATCCGCGATACACAGTCTCTGTTGGCTGTGGTGGAGC 1122
 Db 504 CATTTCTTGATTTGTGAGCATCTGACATCAAGGCGCTTGTATTTGGCAACCGGAGATTCT 445
 Qy 1123 GGAAGAGGGTTAAATATCTGCTTCTTATTTGGGAATCTCATTTGTTGACGGAT----- 1175
 Db 444 GGACAGGATTCAGTATATGCCCATAATTTGGCCACTAGATATCCGACCTCATTTGAAG 385
 Qy 1176 --GGAAGGTAAAGTGGCGCAAAATAATTCACGAATTAATCAAGTGAACCCGCGCATTCGC 1233
 Db 384 GGAATAATCGGGATTATACAGAAACTTATCGCAAGCATGCGCTGGAGACAGAGATTGCC 325
 Qy 1234 GCGAACCGTAATCGGCGTGATCTCTGCGGGGTTTGGCGGTCCAAATCGTGTGATGGAT 1293
 Db 324 GAGGGAAGAGATATATACACCTCCCAAGGGAGGTTTGGTGGCTCAAAACAGAAATCAAGGAC 265
 Qy 1294 TTTTCATGATCTGAAGGAATGACCAATGTTCA 1325
 Db 264 CTCAATGAGCTAAGGATTGGACAAATAGCA 233

RESULT 3

CNS07AGE 906 bp DNA linear GSS 08-JUL-2001
 LOCUS T3 end of clone BC0AA003H11 of library BC0AA from strain CBS 767 of
 DEFINITION Debaromyces hansenii, genomic survey sequence.

ACCESSION AL436548

VERSION AL436548.1 GI:12219961

KEYWORDS GSS.

SOURCE Debaromyces hansenii (anamorph: Candida famata)

ORGANISM Debaromyces hansenii

REFERENCE 1 (bases 1 to 906)
 AUTHORS Souciet J.L., Aigle M., Artiguenave F., Blandin G., Bolotin-Fukuhara M., Bon E., Brottier P., Casaregola S., de-Montigny J., Dujon B., Durrens P., Lepingle A., Llorente B., Malpertuy A., Neveglisse C., Ozier-Kalogeropoulos O., Potier S., Saurin W., Tekala P., Toffano-Nioche C., Wesolowski-Louvel M., Wincker P. and Weissenbach J.

REFERENCE 2 (bases 1 to 906)
 AUTHORS Lepingle A., Casaregola S., Neveglisse C., Bon E., Nguyen H., Artiguenave F., Wincker P. and Gaillardin C.

TITLE Genomic exploration of the hemiascomycetous yeasts: 14.

JOURNAL Debaromyces hansenii var. hansenii

MEDLINE 20584724

PUBMED 11152889

REFERENCE 3 (bases 1 to 906)

AUTHORS Genoscope.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711

PUBMED 11152876

REFERENCE 2 (bases 1 to 906)

AUTHORS Lepingle A., Casaregola S., Neveglisse C., Bon E., Nguyen H., Artiguenave F., Wincker P. and Gaillardin C.

TITLE Genomic exploration of the hemiascomycetous yeasts: 14.

JOURNAL Debaromyces hansenii var. hansenii

MEDLINE 20584724

PUBMED 11152889

REFERENCE 3 (bases 1 to 906)

AUTHORS Genoscope.

TITLE
JOURNAL

COMMENT

Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaromyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES

source

1. .906
 /location=Qualifiers
 /organism="Debaromyces hansenii"
 /mol_type="genomic DNA"
 /strain="CBS 767"
 /variety="hansenii"
 /db_xref="taxon:4959"
 /clone="BC0AA003H11"
 /clone_lib="BC0AA"
 /notes="end : T3"
 /notes="end : T3"

misc_feature

<15..>716
 /notes="similar to O43029 [PUTATIVE FRUCTOSYL AMINO ACID
 OXIDASE]"
 /evidence=not_experimental

ORIGIN

Query Match 7.0%; Score 99.2; DB 9; Length 906;
 Best Local Similarity 50.8%; Pred. No. 3.3e-19;
 Matches 335; Conservative 0; Mismatches 298; Indels 27; Gaps 3;
 Qy 498 TGGCTGGGCACATGCAAGGAATGCTTAGTGGCAGCAGCAGCGAAGCACAGCGCATGGG 557
 Db 14 TGGGTGGGCATTTGCTAGGTTGGCATTAGAAGACTGTTGAAGAAATGCCGTAATAGG 73
 Qy 558 TGTAAATTTGTTACTGGCACCCCGCAGGTCGTGTAGTCACTTAATCTTTGAAAAATAA 617
 Db 74 GGCTAAATTTGTTATTGATTCTGCTGAAGAACTTCTCTTCA-----GTGAAGATGG 124
 Qy 618 CGATGTAAAGGTGCCGTTAGCGGCGATGCGCAAAATTTGGAGAGCGGAACGTACATTCCT 677
 Db 125 AGCATGTGTGGTGGTTCACACTTCGAATGGAACATTTATTGAAGCTGATGAACATTTAT 184
 Qy 678 GTGTCTGGGCTAGCGCGGTGAGTTCCTAGATTTCAAGAAATCAACTTTCGACCAACCGC 737
 Db 185 TTGTGCTGGTGCATAATTCATTCAGTTTTTTGAACCTTTGAGCAGCACTTTTGGCTAATG 244
 Qy 738 TTGGACCTCTGGTACACATTCGCTTAAACCGGAAGAACGTCGCTTGTACAAAAATATACC 797
 Db 245 CTATACCTTTGGACATATCAAACTCACTGATGATGAAGTCTTTTGTGAAGGGGATGCC 304
 Qy 798 GGTATCTTTAAACATCGAAGCGGGGTTTCTTTTGAACCGCATGAGGAGCGCGGTGAGAT 857
 Db 305 AGTTGTTTAAATCTAGATGGAGGTTTGTATTGGAACCGGACTTAA---CAATGAGAT 361
 Qy 858 TAAATATGCGATGAACACACCGGCGGTACACAAATATGTCAGAGTCCAGACGCGACGAT 917
 Db 362 TAAATCTGTAATGAGTTTCCAGGTATGTCAACATCGTAATGGAAGATTTCAG----- 414
 Qy 918 GATGAGCATTCCTGTTGAAAAAACCCAGATTTCAAAAAGAACCGGAACCGCGCTTCGGGC 977
 Db 415 -----TGCCTTCATTTAAGGATTCAATCCCAAGAGGCTGAAGATCAATGAGAGC 466
 Qy 978 CTGTGTAAGAGACAAATGCCAGCTGGCAGACCGCTCCATTCAGTCTGCACCGATTG 1037
 Db 467 ATTTTAAAGACAAAGTTTTCCTCCGAAATTTGCTGAAAGGGAGTTTTCATTAGTAGAATTG 526
 Qy 1038 CTGTGTGCGGATACCGCGAATCGCAATTCCTGTATAGATCGACATCGGAGTACCCAG 1097
 Db 527 CTGTGCACAGACACTCCAGATCGTCATTTTGTGAGCATCCGGGGCATATAAA 586


```

Db      277 ATACAGCACATTACATCTATTCGCAATTCATCTCGATTGTATGAGGGTAGATTGGA 218
Qy      1191 GCAAAAATTACGAATTAATCAAGTGGAAACCCGGAATTCGGCGGAAACCGTAACTGGCG 1250
Db      217 AGAGAGATTTCGCAAGTTTGGACATGGAGACCGGAGAAAGTTTACTGAGTTTGGGGTAA 158
Qy      1251 TGATACTCTGGGGGTTTTCGGCGTCCAAATCGTGTGATGATTTTC 1297
Db      157 CGATCCTCTAGATCGATTTCGTCTGACGATCATCATGATCTGC 111

RESULT 6
LOCUS   CN241854
DEFINITION EST007722 Mycelium and yeast cells from Paracoccidioides brasiliensis Paracoccidioides brasiliensis cDNA, mRNA sequence.
ACCESSION CN241854
VERSION   CN241854.1 GI:46345598
KEYWORDS  EST.
ORGANISM Paracoccidioides brasiliensis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Paracoccidioides.
REFERENCE 1 (bases 1 to 677)
AUTHORS  Felipe, M.S.S., Carvalho, M.J.A., Andrade, R.V., Arraes, F.B.M., Simoes, I.C., Andrade, E.V., Maranhao, A.Q., Torres, F.A.G., Jesuino, R.S.A., Kyaw, C.M., Moraes, L.M.P., Nicola, A., Pereira, M., Silva-Pereira, I., Anjos, D.A.S., Sandes, E.F.O., Inoue, M.K., Walter, M.E.M.T., Soares, C.M.A. and Brígido, M.M.
TITLE    Metabolic features of Paracoccidioides brasiliensis cell differentiation as accessed by transcriptome analysis
JOURNAL  Unpublished (2004)
COMMENT  Contact: Felipe MSS
          Laboratory of Molecular Biology
          Institute of Biology - University of Brasilia
          Campus Universitario, Asa Norte, Brasilia, DF 70910-900, BRA
          Tel: 55 61 307 2423
          Fax: 55 61 349 8411
          Email: msueli@unb.br
          Seq primer: T7 Sequencing primer.

FEATURES             source
    source
        1..677
            /organism="Paracoccidioides brasiliensis"
            /mol_type="mRNA"
            /strain="Pb01"
            /db_xref="taxon:121759"
            /clone_lib="Mycelium and yeast cells from Paracoccidioides brasiliensis"
            /note="Pb Lambda Zap Express Library"

ORIGIN
Query Match      4.7%; Score 66.2; DB 7; Length 677;
Best Local Similarity 51.2%; Pred. No. 7.7e-09;
Matches 155; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy      933 CGAAAAAACCCAGATTCCAAAAGAACCGGCGTTTCGGGCGCTGCTGAAAGAGAC 992
Db      243 CGGAGATGTCCGATCCCGATTCAGGGGAGGAGCATGTAGAGCCGCGCTGAGAGAGAT 302
Qy      993 AATGCCCCAGCTGGCAGACCGTCCATTCAGCTTCGACGCGATTCGTTGTCGCCGATAC 1052
Db      303 GCTGCCCGCATTTTCAGAGCGGGCGTTTGTGAACACCGCGAGTGTGTTGGTATCCGGATAC 362
Qy      1053 CGCGAATCGGGAATTCCTGTATAGATCGACATCGCAGTACACAGCTCTGTGTGTGGCGTG 1112
Db      363 ACCACAGGGCGACTTCTAGTAACATACACACACTACCGGGCGCTCTCTCTCGCAAC 422
Qy      1113 TGGTGGCAGCGGAGAGGGTTTAAATATCTGCCCTTCTTATTTGGGAATCTCAATGTTGACGC 1172
Db      423 CGGTGGCAGCGGCGACGGGTTTAAATTCCTCCCGCTCGGGGAGAAAGATCGTCGATGC 482
Qy      1173 GATGGAAGGTAAGTGGCCGCAAAAATTCAGGAATTAATCAAGTGGAAACCGGACATTCG 1232

```

```

Db      483 GATACAGGACGGTTTGATCTGAGTTGCAGAGTATATGCGGTGGACTAGTCTACTTC 542
Qy      1233 GGC 1235
Db      543 GTC 545

RESULT 7
LOCUS   Bi187570
DEFINITION Bi187570.1
ACCESSION Bi187570
VERSION   Bi187570.1 GI:14661249
KEYWORDS  EST.
SOURCE    Fusarium sporotrichioides
ORGANISM  Fusarium sporotrichioides
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Fusarium.
REFERENCE 1 (bases 1 to 414)
AUTHORS  Ren, Q., Tag, A., Peplow, A., Lai, H., Kupfer, C., Peterson, A., Beremand, M. and Roe, B.
TITLE    Analysis of a Fusarium sporotrichioides EST database
JOURNAL  Unpublished (2001)
COMMENT  Other_ESTs: a4902fs.fl
          Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
          Department of Chemistry and Biochemistry
          Advanced Center for Genome Technology, University of Oklahoma
          620 Parrington Oval, Norman, OK 73019, USA
          Tel: 405 325 4912
          Fax: 405 325 7762
          Email: broe@ou.edu
          Contact Dr. Marian Beremand regarding clone availability Included
          is the best homolog from a blastx search of Genbank nr 04-09-01
          217 3e-29 gi|2661130|gb|AAB882 (AF035700) fructosyl
          amine:oxygenoxidoredu
          Seq primer: T3.

FEATURES             source
    source
        1..414
            /organism="Fusarium sporotrichioides"
            /mol_type="mRNA"
            /strain="Tri 10"
            /db_xref="taxon:5514"
            /clone="a4902fs"
            /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed cDNA library"
            /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript; 3' end of cDNA cloned into XhoI site of pBluescript"

ORIGIN
Query Match      4.6%; Score 65.2; DB 4; Length 414;
Best Local Similarity 54.7%; Pred. No. 1.4e-08;
Matches 176; Conservative 0; Mismatches 138; Indels 8; Gaps 2;

Qy      31 GCTCTCGCGCTCACTAAGTCATCTCTCTGATCTGTGTCGCGGACTTGGGGCACC 90
Db      12 GCTTCATCTCTTACCAAGAGTCCAAATCTCATCGTTGGAGTGGGACATGGGATGT 71
Qy      91 TCAACGGCTCTGCACCTTCGCGCGCGGATATACCAACGTTTACCGTGTGGACCCCTAT 150
Db      72 TCCACAGCTCTTCATCTAGCTCGTCGAGGCTATACCAAGTCACTTCTCGA--TGTTA 129
Qy      151 CTGTCTCCTAGCCCATCTCCGCGGAAACGAGTGAACAAAGTCATTAG-----CAGT 204
Db      130 ACCATCCCTTACCACAAATCTCTCGGGCCATGATGTCAACAAGATTGCAAGTGGATTGAAT 189
Qy      205 GGCCAAATATTTCGAATTAACAAGACGAATCGAAGTGAATGAGATCTTTGGCGGAAGAGCG 264
Db      190 GTCCCGATTAATAAGAGACGATGAAGATGCAATCTGGCAATCCCTCACCCTACCTACGCCCA 249

```

QY 265 TTTAACGGTTGGAAGAACGACCGCTTTTCAACCGTATTATCATGATACGGCCTGCTG 324
Db |||||
250 GCTCAGGCTGGCTACATGATCCAGTCTTCAAAACCATATACCAACGACAGGCTATGTC 309
QY 325 ATGCTCTGCTTCTCGCAGGAGG 346
Db |||||
310 ATCTCTGCTTCAACGCGCAAAGG 331

RESULT 8
AJ636269
LOCUS
DEFINITION AJ636269 Mga Mycosphaerella graminicola cDNA clone mgal1757f, mRNA sequence.
ACCESSION AJ636269
VERSION AJ636269.1 GI:47029326
KEYWORDS EST.
SOURCE Mycosphaerella graminicola
ORGANISM Mycosphaerella graminicola
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes et Chaetothyriomycetes incertae sedis; Mycosphaerellaceae; Mycosphaerella.
1 (bases 1 to 599)
Contact: Keon J
Keon, J.P.R., Bailey, A.M. and Hargreaves, J.A.
A group of expressed cDNA sequences from the wheat fungal leaf blotch pathogen, Mycosphaerella graminicola (Septoria tritici) Fungal Genet. Biol. 29 (2), 118-133 (2000)
20374020
JOURNAL
MEDLINE
PUBMED 10919380
COMMENT Contact: Keon J
Plant Pathogen Interactions Division,
Rothamsted Research,
Harpenden, Herts, UNITED KINGDOM
Tel: +44(0)1582 763133
Fax: +44(0)1582 760981
Email: john.keon@bbsrc.ac.uk
Insert Length: 800 Std Error: 100.00
Seq primer: M13 reverse.

FEATURES

source

Location/Qualifiers
1..599
/organism="Mycosphaerella graminicola"
/mol_type="mRNA"
/strain="Strit"
/db_xref="taxon:54734"
/clone="mgal1757f"
/clone_lib="Mga"
/notes="Vector: pSPORT1; Library constructed from cultures grown in Czapek Dox for 3 days then Glucose Yeast Nitrogen Base for 1 day, shift of nitrogen source from nitrate to ammonium ions"

ORIGIN

Query Match 4.6%; Score 64.6; DB 1; Length 599;
Best Local Similarity 53.8%; Pred. No. 2.4e-08;
Matches 133; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
QY 937 AAACCCAGATTCCAAAGACCGCAACCGCGTTTCGGGCCCTGCTGAAGACAAATG 996
Db |||||
217 AACATGCAATCCACCGGAAGAGAGCGTTCGCTCGCGTCTCTCGCAAAAGCCATA 276
QY 997 CCCAGCTGGCAGACCGTCCATTTCAGCTTCGCGCGCAATTTGCTGTGTGCCGATACCGG 1056
Db |||||
277 CCTCGATGGCGACCGACCACTGGGCCAACACCGGATATGCTGTATACAGGACCGTC 336
QY 1057 AATCGCGAATCTCTGATAGATCGACATCCGCACTACACAGTCTTGTGTTGGCTGTGGT 1116
Db |||||
337 TCTGGAGACTGGTTGATCATCAGCTACCATCCGAAGTATTCCGGGACTCTTCGTCGCGACGGC 396
QY 1117 GCAGCGGAAGAGGGTTTAATATCTGCTTCTATTGGGATCTCATTTGTTGCGGATG 1176
Db |||||
397 GGCTCGGACATGATGATAAATCTTTGCCGATCATTTGGAGGGAGAAATTTGGATGTGATT 456

QY 1177 GAAGGTA 1183
Db |||||
457 GAGGGCA 463

RESULT 9

BM869701

LOCUS

DEFINITION

BM869701 628 bp mRNA linear EST 06-MAY-2003
mgns007xF15f.b Magnaporthe grisea NS Uni-Zap XR Library Magnaporthe grisea cDNA clone mgns007xF15 5', mRNA sequence.

ACCESSION

BM869701

VERSION

BM869701.2

KEYWORDS

EST.

SOURCE

Mgna

ORGANISM

Magnaporthe grisea

REFERENCE

AUTHORS

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.

TITLE

1 (bases 1 to 628)

JOURNAL

COMMENT

On Mar 7, 2002 this sequence version replaced gi:19237383.
Contact: Ebbolte DJ
Department of Plant Pathology & Microbiology
Texas A&M University
Peterson Bldg. MS2132, College Station, TX 77843-2132, USA
Tel: 979 845 4831
Fax: 979 845 6483
Email: d-ebbolte@tamu.edu
Chromatogram file of this sequence is available, see contact person; Best nr hit (April. 22, 2003) gb|EAA30984.1| hypothetical protein [Neurospora crassa] 194 6e-49
PCR Primers
FORWARD: T3 primer
BACKWARD: T7 primer
Plate: mgns007 row: F column: 15
Seq primer: T3.

FEATURES

source

Location/Qualifiers

1..628
/organism="Magnaporthe grisea"
/mol_type="mRNA"
/strain="Guy11"
/db_xref="taxon:148305"
/clone="mgns007xF15"
/sex="Mat1-2 hermaphrodite"
/cell_type="mycelium"
/clone_lib="Magnaporthe grisea NS Uni-Zap XR Library"
/notes="Vector: pBluescriptSK-; Site 1: EcoRI; Site 2: XhoI; Unidirectional cloning. EcoRI side has T3 primer and predominantly 5' reads. T7 primer on XhoI side of insert. Nitrogen starvation library. Cells were inoculated into minimal medium and grown for two days with shaking (150 rpm) at room temperature. Culture was harvested, blended, inoculated into minimal medium as above for 24 h. Cells were harvested, washed with water and inoculated into minimal medium base lacking nitrogen source for 6 h. Sequences were processed by one of two methods. Where a full-length alignment to the M. grisea genome sequence was available, the EST sequence was trimmed according to the alignment, otherwise sequence quality was assessed using phredPhrap version 991019 and trimmed according to phd files (0.05) and for vector seqs."

ORIGIN

Query Match 4.6%; Score 64.6; DB 4; Length 628;
Best Local Similarity 56.3%; Pred. No. 2.4e-08;
Matches 121; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 992 CAATGCCCGCTGGCAGACCGCTCCATTTCAGCTTCGACCGCATTTGCTGTGTCGCATA 1051
Db |||||
217 CGATCTCTCGATTTCATGACCGCGCCCTTCACTTCAACCGCGCTATGCTGTACACCGACA 276

Qy	1052	CCGCGAATCGCGAATTCCTGATAGATCGACATCTCGCAGTAGTACCAGATCTCTGTGTGGGCT	1111
Db	277	CAAAAGTCGGCGCATTTTCATCATCGATTACCATCTCGAGTCGGGAAGGTCTTTTCTTCGCCCA	336
Qy	1112	GTGTGTGGAGCGGAAGAGGTTTTAAATATCTGCCTTCTATTGGGAATCTCATTTGTTGCAAG	1171
Db	337	CTGTTGTGTACGGCCATGGTTTCAAGTCTTTCGCGTCTTTGGAGAACGGATTGCCGACT	396
Qy	1172	CGATGGAAGGTAAAGTCCGCCGAAAAAATTCACGAA	1206
Db	397	GTATCGCATATGATCGACCGCTCAGAAATTTGCGGAA	431

RESULT 10
 CF188389/c
 LOCUS
 DEFINITION
 459 bp mRNA linear EST 08-AUG-2003
 k2d07j2.f1 Cryptococcus neoformans strain B3501 Cryptococcus
 neoformans var. neoformans cDNA clone k2d07j2.3, mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 CF188389.1 GI:33510258
 EST.
 Cryptococcus neoformans var. neoformans (Filobasidiella neoformans
 var. neoformans)
 Cryptococcus neoformans var. neoformans
 Karyota; Fungi; Basidiomycota; Hymenomycetes;
 Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae;
 Filobasidiella.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 1 (bases 1 to 459)
 Kupfer D.M., Drabenstot, S.D., Buchanan, K.L., Lai, H., Dyer, D.W.,
 Roe, B.A. and Murphy, J.W.
 Comparison of highly conserved intronic and exonic elements
 associated with splicing among five diverse fungal organisms
 Unpublished (2003)
 Other ESTs: k2d07j2.r1
 Contact: Murphy, JW
 Department of Microbiology and Immunology
 University of Oklahoma Health Sciences Center
 Oklahoma City, OK 73190, USA
 Tel: 405-271-2133 ex2133
 Email: juneann-murphy@ouhsc.edu
 This clone is available from the Fungal Genetics Stock Center,
 phone 913-588-7044, <http://www.fgsc.net/> Contact Dr. Bruce Roe
 (broe@ou.edu, www.genome.ou.edu) for sequencing questions Contact
 Dr. Juneann Murphy (juneann-murphy@ouhsc.edu) for library
 information.

Qy	1065	ATTCTGATAGATCGACATCCGCGAGTACACAGTCTTGTGTGGCGTGTGGTCGAGCG	1124
Db	321	CTGGCTACTATCGCGACATCCCGAGTGGAAAGTTGGTTGGCTACGGAGATATGTG	262
Qy	1125	AAGAGGGTTTAAATATCTGCCTTCTATATCGGAATCTCATTTGTGACGGATGGAAGGTAA	1184
Db	261	GCACACCTTTAAATGCTGCCTGTGTGGCGCCAAAGTTGCTGACCTCATCGAAGCAA	202
Qy	1185	AGTCCGCGAAAAATTCACGAATTAATCAAGTGGAAACCCGG	1225
Db	201	AATGTCGAGCAGAGAAAGATGCTCTTGGAGGTCGAGACCTG	161

RESULT	11
BG465351	
LOCUS	
DEFINITION	422 bp mRNA linear EST 20-MAR-2001 EM1_75_C06_g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION	BG465351
VERSION	BG465351.1 GI:13394327
KEYWORDS	EST.
SOURCE	Sorghum bicolor (sorghum)
ORGANISM	Sorghum bicolor
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Focaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 422) Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H. An EST database from Sorghum: developing embryos Unpublished (2000) Contact: Cordonnier-Pratt MM
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

FEATURES	Location/Qualifiers
source	1. .422
	/organism="Sorghum bicolor"
	/mol_type="mRNA"
	/db_xref="taxon:4558"
	/clone_lib="Embryo 1 (EM1)"
	/notes="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

```
QY 1124 GAAGAGGGTTTAATATCTGCCTTCTATTGGGAATCTCAATGTTGACGCGATGGAGGTA 1183
Db 200 GGCATGGCTACAGATTTCGTGCCAGATTGGCGAGAGGATCGTGGATGTAATGAAGCGCC 259
QY 1184 AAGTCCGCGAATAATTCAGGAATTAATCAAGTGGACCGGACATTTGGCGGAACCGTA 1243
Db 260 AACCACGGATGACTTGGGTACTGAAATTCAGAGAAGATGGTTCATGGCGGACGAGAAT 319
QY 1244 ACTGGCGTGATCTCTGGGCGGTTTGGCGGTCCAAATCGTGTGATGGATTTTCATGATG 1303
Db 320 ACCACTTTCATCATCTGCAAGCACTGGAGAGTGGACGAAGGATGATGATTGTTG 379
QY 1304 TGAAGGAATGGACCAATGTTCACTA 1328
Db 380 AGGAAGACGAGAGGAGGGCGGAA 404

RESULT 12
CF871025 758 bp mRNA linear EST 31-OCT-2003
LOCUS
DEFINITION
Hypocrea jecorina cDNA clone trico25xol7, mRNA sequence.
ACCESSION
CF871025
VERSION
CF871025.1 GI:38125707
KEYWORDS
EST.
SOURCE
Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM
Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
Diener, S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,
Teunissen, P.J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K.,
Ward, M. and Dean, R.A.
Characterization of the protein processing and secretion pathways
in a comprehensive set of expressed sequence tags from Trichoderma
reesei.
JOURNAL
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
COMMENT
Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: LT-F1 primer.
FEATURES
source
1..758
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trico25xol7"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/notes="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."
ORIGIN
Query Match 3.9%; Score 55.2; DB 7; Length 758;
Best Local Similarity 60.8%; Pred. No. 2.3e-05;
Matches 90; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 53 CATCTCTCTGATCGTTGGTGGCGGACCTTGGGGACCTCAACGGCTCTGCACCTGCGCG 112
Db 346 CGTCAGTTCTGATTGTTCGGTCTGTTAACTTCGGCGCGCAACGGCTTTGAGTCTGGCCA 405
QY 113 GCCGCGGATATACCAAGTTACCGTCTGGACCCCTATCTGTCCTTAGCGCCATCTCCG 172
Db 406 GAAGGGGAGGGTTCAAGGTCAACATCATTTGACACGGCCCCCTTCCCCAACCCAGGCGCG 465
QY 173 CCGGAACACGAGTGAACAAAGTCATTAG 200
```

```
Db 466 CCTCCACGACATCAACAGATCGTTCG 493

RESULT 13
CB901239 816 bp mRNA linear EST 02-JUL-2003
LOCUS
DEFINITION
Hypocrea jecorina cDNA clone trico25xol7, mRNA sequence.
ACCESSION
CB901239
VERSION
CB901239.1 GI:30115897
KEYWORDS
EST.
SOURCE
Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM
Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)
22803314
PUBMED
12788920
COMMENT
Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: pforeman@genencor.com
Seq primer: LT-F1 primer.
FEATURES
source
1..816
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trico25xol7"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/notes="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."
ORIGIN
Query Match 3.9%; Score 55.2; DB 6; Length 816;
Best Local Similarity 60.8%; Pred. No. 2.4e-05;
Matches 90; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 53 CATCTCTCTGATCGTTGGTGGCGGACCTTGGGGACCTCAACGGCTCTGCACCTGCGCG 112
Db 404 CGTCAGTTCTGATTGTTCGGTCTGTTAACTTCGGCGCGCAACGGCTTTGAGTCTGGCCA 463
QY 113 GCCGCGGATATACCAACGGTTACCGTCTGGACCCCTATCTGTCCTTAGCGCCATCTCCG 172
Db 464 GAAGGGGAGGGTTCAAGGTCAACATCATTTGACACGGCCCCCTTCCCCAACCCAGGCGCG 523
QY 173 CCGGAACACGAGTGAACAAAGTCATTAG 200
Db 524 CCTCCACGACATCAACAGATCGTTCG 551

RESULT 14
CB901049 822 bp mRNA linear EST 02-JUL-2003
LOCUS
DEFINITION
Hypocrea jecorina cDNA clone trico25xe08, mRNA sequence.
ACCESSION
CB901049
VERSION
CB901049.1 GI:30115707
KEYWORDS
EST.
SOURCE
Hypocrea jecorina (anamorph: Trichoderma reesei)
```

ORGANISM *Hypocrea jecorina*
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE 1 (bases 1 to 822)
 AUTHORS Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
 Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
 Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
 Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
 TITLE Transcriptional regulation of biomass-degrading enzymes in the
 filamentous fungus *Trichoderma reesei*
 J. Biol. Chem. 278 (34), 31988-31997 (2003)

JOURNAL 22803314
 MEDLINE 12788920
 PUBMED
 COMMENT Contact: Pamela K. Foreman
 Genencor Intl.
 925 Page Mill Road, Palo Alto, CA 94304, USA
 Tel: (650) 846-7635
 Fax: (650) 621-7817
 Email: Pforeman@genencor.com
 Seq primer: LT-F1 primer.
 Location/Qualifiers
 source 1..822
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="trix025xe08"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 3 april"
 /note="vector: pREPxy; Site_1: Not I/Sal I; Mycelial
 culture grown from 24 hrs to 6 days with varying Carbon
 and Nitrogen sources and concentrations."

ORIGIN
 Query Match 3.9%; Score 55.2; DB 6; Length 822;
 Best Local Similarity 60.8%; Pred. No. 2.4e-05;
 Matches 90; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 53 CATCTCTCTGATCGTTGGTCCGGGACCTCGGGGCACCTCAACGGCTCTGCACCTCGCGC 112
 Db 420 CGTCAGTTCTGATTGTCGGTCTGTTAACTTCGGCGCGCAACGGCTTTGAGTCTGGCCA 479

Qy 113 GCGCGGATATACCAAGTTACCTGCTGGACCCCTATCTCTGTCCTTAGCGCCATCTCCG 172
 Db 480 GAAGGGGAGGGTTCAAGTCAACCATCTGACACGGCCCTCTCCCAACCCAGGGCGC 539

Qy 173 CCGGAACGAGCTGAACAAAGTCATTAG 200
 Db 540 CTTCCACGACATCAACAAAGATCTTGC 567

RESULT 15
 CF870843
 LOCUS CF870843 822 bp mRNA linear EST 31-OCT-2003
 DEFINITION *Hypocrea jecorina* mycelial culture, Version 6 October 2003
 ACCESSION CF870843
 VERSION CF870843.1 GI:38125525
 KEYWORDS EST.
 SOURCE *Hypocrea jecorina* (anamorph: *Trichoderma reesei*)
 ORGANISM *Hypocrea jecorina*
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE 1 (bases 1 to 822)
 AUTHORS Diener,S.E., Dunn-Coleman,N., Foreman,P., Houfek,T.D.,
 Teunissen,P.J.M., van Solingen,P., Dankmeyer,L., Mitchell,T.K.,
 Ward,M. and Dean,R.A.
 TITLE Characterization of the protein processing and secretion pathways
 in a comprehensive set of expressed sequence tags from *Trichoderma reesei*

JOURNAL FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
 COMMENT Contact: Ralph A. Dean
 Fungal Genomics Laboratory

THIS PAGE BLANK (USPTO)